

B.Z.VULIKH

*A Brief Course
in the Theory
of Functions of
a Real Variable*

The author of this book, Prof. B. Z. Vulikh, D. Sc. (Math.) is well known for his original research in the field of functional analysis, particularly in the theory of partially ordered spaces and for his pedagogical activity as head of the Chair of Mathematical Analysis at the Leningrad State University.

His publications include **Functional Analysis in Partially Ordered Spaces** (co-author), **An Introduction to Functional Analysis**, and **Introduction to the Theory of Partially Ordered Spaces**.

The present book is an introduction for university students to the theory of measure and of integrals, primarily of the Lebesgue integral. It can be used as a textbook either for a separate course in this subject or as part of the general university course in mathematical analysis.



HENRIQUE LAZARI

Ganhei do Henrique!

Dia 02/05/2016



THE NEW YORK PUBLIC LIBRARY

ASTOR LENOX TILDEN FOUNDATION

155 E. 42ND STREET, NEW YORK 17, N.Y.

1964



Б. З. Вулих

**КРАТКИЙ КУРС
ТЕОРИИ ФУНКЦИИ
Вещественной переменной**

Введение в теорию интеграла

МОСКВА «НАУКА»

Marta C. Gadotti

B. Z. VULIKH

maio/
2016

A BRIEF COURSE
IN THE THEORY OF FUNCTIONS
OF A REAL VARIABLE

(An introduction
to the theory
of the integral)

Translated from the Russian

by I. G. Volosova

Mir
Publishers
Moscow

First published 1976
Revised from the 1973 Russian edition

(An introduction
to the theory
of the integral)

На английском языке

© Издательство «Наука», 1973

© English translation, Mir Publishers, 1976

Contents

Preface	9
CHAPTER I. Fundamentals of Set Theory	11
§ 1. Basic Operations on Sets	11
§ 2. Some Auxiliary Relations	17
§ 3. The Power of a Set	18
§ 4. Countable Sets	21
§ 5. The Power of the Continuum	26
§ 6. Comparison of Powers	32
§ 7. Rings, Semirings and Algebras of Sets	34
CHAPTER II. Point Sets in a Euclidean Space	39
§ 1. n -dimensional Euclidean Space	39
§ 2. Limit Points	42
§ 3. Closed and Open Sets	46
§ 4. The Structure of an Open Set on the Line	49
§ 5. The Structure of an Open Set in the n -dimensional Space R_n	51
§ 6. Finite Covering Theorem	53
§ 7. Continuous Functions	55
§ 8. Borel Sets	57
CHAPTER III. Metric Spaces	59
§ 1. The Definition of a Metric Space	59
§ 2. Convergence in a Metric Space	63
§ 3. Closed and Open Sets	66
§ 4. Complete Metric Spaces	68
§ 5. Separable Spaces	72
§ 6. Normed Spaces	74
§ 7. Linear Functionals in Normed Spaces	77
CHAPTER IV. Measure in Abstract Sets	83
§ 1. Additive Set Functions	83
§ 2. Definition of a Measure. Properties of Measures	86
§ 3. Outer Measure	88

§ 4. Standard Extension of Measure	94
§ 5. The Uniqueness of the Extension of Measure	97
CHAPTER V. The Lebesgue Measure in a Euclidean Space	101
§ 1. n -dimensional Parallelepipeds	101
§ 2. The Volume of a Parallelepiped	104
§ 3. Semiring of Cells	109
§ 4. Representation of an Open Set with the Aid of Cells	112
§ 5. Measurable Sets	115
CHAPTER VI. Measurable Functions	121
§ 1. The Definition of a Measurable Function	121
§ 2. Arithmetical Operations on Measurable Functions	126
§ 3. Limit in the Class of Measurable Functions	128
§ 4. Equivalent Functions. Convergence Almost Everywhere	132
§ 5. Convergence in Measure	135
§ 6. Convergence Regulator. Theorems of D. F. Egorov, N. N. Luzin and R. M. Fréchet	141
CHAPTER VII. The Lebesgue Integral of a Bounded Function	150
§ 1. The Definition of the Lebesgue Integral	150
§ 2. The Simplest Properties of the Integral	156
§ 3. Passage to the Limit in the Lebesgue Integral	165
§ 4. The Space S of Measurable Functions	167
CHAPTER VIII. Summable Functions	180
§ 1. The Definition of a Summable Function	180
§ 2. Lemmas on Integrals of Nonnegative Functions	183
§ 3. Generalization of the Simplest Properties of Integrals	190
§ 4. Passage to the Limit under the Integral Sign	200
§ 5. The Space L of Summable Functions	213
§ 6. Geometrical Interpretation of Lebesgue's Integral in Euclidean Space	220
§ 7. Iterated Integrals	225
§ 8. Product Measure	233
CHAPTER IX. Square Summable Functions	237
§ 1. The Space L^2	237
§ 2. Scalar Product	242
§ 3. Orthogonal Series	246
§ 4. Linear Functionals on L^2	253

CHAPTER X. The Space L^p	257
§ 1. The Definition of the Space L^p . Basic Inequalities	257
§ 2. Mean Value of a Function	262
§ 3. An Inequality for Iterated Integrals	265
§ 4. Linear Functionals on L^p	268
CHAPTER XI. The Radon Integral	271
§ 1. Variations of Additive Set Functions	271
§ 2. The Radon Integral	281
§ 3. Integration with Respect to a Finitely Additive Set Function	287
§ 4. Linear Functionals on the Space of Bounded Functions	289
§ 5. The Stieltjes Integral on the Line	294
§ 6. Distribution Functions on the Line	299
CHAPTER XII. Absolutely Continuous Set Functions	306
§ 1. Definition of an Absolutely Continuous Set Function	306
§ 2. The Radon-Nikodym Theorem	308
§ 3. The Hahn Decomposition	313
CHAPTER XIII. Indefinite Lebesgue Integral	315
§ 1. Absolutely Continuous Point Functions	315
§ 2. Properties of Functions Representable in the Form of an Integral	319
§ 3. Differentiation of Continuous Monotone Functions	321
§ 4. Differentiation of Discontinuous Monotone Functions	331
§ 5. Differentiation of the Integral	337
§ 6. Criterion of F. Riesz	343
§ 7. Density Points of a Set on the Line	346
Bibliography	351
Name Index	353
Subject Index	354

1 The first of these is the fact that the
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

Preface

The aim of this course is to give a brief representation of those aspects of the theory of functions of a real variable which are most widely applied in other divisions of mathematics, in particular, in the theory of partial differential equations and in the theory of integral equations. Therefore the subject matter of the book is primarily concerned with the measure theory and the Lebesgue integral. The set theory is only slightly touched upon to provide some fundamentals needed for the understanding of the basic questions treated in the book.

The modern theory of functions of a real variable is closely interrelated with functional analysis. The former creates the basis for studying many concrete function spaces and the latter provides a more general modern approach to the problems of the theory of functions. That is why this book also includes some elements of functional analysis, namely, the basic notions of the general theory of metric spaces and the theory of normed function spaces, which are then applied to studying various classes of real functions from the point of view of functional analysis.

Some branches of the theory of functions of a real variable, for instance, the so-called descriptive theory of functions, are not included in the book. The bibliography at the end of the book includes some textbooks and monographs which are partially or completely devoted to questions related to the theory of functions of a real variable. We suggest that the reader pay particular attention to the book by I. P. Natanson, *Theory of Functions of a Real Variable* (the English translation: New York, V. 1, 1955; V. 2, 1960), which not only treats of many important topics not included in the present course but also contains a wide variety of problems whose solution is useful exercise for the reader.

The reader is supposed to be familiar with a comprehensive course of mathematical analysis studied, for instance, at mathematical departments of universities.

For the sake of convenience, the theorems and lemmas are independently numbered in a consecutive order within each section of the book. For instance, "Theorem VII.3.4" means "the fourth theorem in Section 3 of Chapter VII". The formulas are numbered in succession within each chapter. When a formula is referred to in the chapter in which it first appears we only give its number in this chapter but if it is mentioned in some other chapter we indicate two numbers: the number of the chapter it originally belongs to and the number of the formula itself within that chapter. In references to a section of the book we also give two numbers indicating, respectively, the number of the corresponding chapter and the number of the section in this chapter.

The author expresses his warmest gratitude to his colleagues at Leningrad university D. A. Vladimirov, B. M. Markarov and G. I. Natanson for helpful comments and valuable advice*.

B. Z. Vulikh

* The present translation incorporates suggestions made by the author.—Tr.

CHAPTER I

Fundamentals of Set Theory

§ 1. Basic Operations on Sets

In the present chapter we shall consider some fundamentals of the set theory which will be used in this course*. The concept of a set is too general to be formally defined. On the other hand, we can easily do without such a definition since we habitually deal with sets both in everyday life and in science. For instance it is quite clear what is meant by the set of books standing in the book-case or what is meant by a set of people living in a certain street or in a certain town. In courses of classical mathematical analysis we deal with the set of all real numbers and with its various parts. Instead of the term "set" we shall often synonymously use the equivalent terms "class", "family", "collection" (and, sometimes, "system").

By *elements* or *members* of a set we mean the "things" or mathematical objects the given set consists of. If a set is denoted by A and its arbitrary element by x we write

$$A = \{x\}$$

If it is necessary to indicate that an object a is one of the elements of a set A we use the so-called *inclusion relation* (*membership relation*) $\in$ and write

$$a \in A$$

(read " a belongs to A " or " a is a member of A " or " a is contained in A "). If an object a is not an element of a set A we write $a \notin A$ (read " a does not belong to A " or " a is not a member of A " or " a is not contained in A ").

* The set theory as an independent branch of mathematics was founded by the German mathematician Georg Cantor (1845-1918) in the second half of the 19th century.

For instance, let A be the set of all positive even integers. Since the general form of any positive even integer is $2n$ (where n is a natural number) we can write $A = \{2n\}$. For example, we have $24 \in A$ while $25 \notin A$.

Let us consider two sets A and B . If every element of the set B is also contained in the set A the set B is called a *subset* of the set A . Such a situation is expressed with the aid of another inclusion relation $\subset$, and we write

$$B \subset A$$

(read " B is a subset of A " or " B is contained in A " or " A contains B ").

For example, the set N of all natural numbers is a subset of the set R of all rational numbers:

$$N \subset R$$

It should be noted that the relation $B \subset A$ does not exclude the coincidence of B and A , that is every set A is one of its subsets.

Inclusion relations are sometimes also written as

$$A \ni a \text{ (instead of } a \in A)$$

and

$$A \supset B \text{ (instead of } B \subset A)$$

Two sets are said to be *equal* (written $A = B$) if they coincide, that is if they consist of precisely the same elements.

It is obvious that if, given two sets A and B , the inclusion relations $B \subset A$ and $A \subset B$ hold simultaneously this exactly means that $A = B$. If $B \subset A$ but $B \neq A$ we call B a *proper subset* of A .

If x and y are elements of some sets the sign of equality between them (i.e. $x = y$) is also used to indicate their coincidence.

A set is usually specified by a certain criterion indicating whether or not a given object is contained in it. A finite set, that is a set consisting of a finite number of elements, can be specified by enumerating all its elements. However, when dealing with a set we sometimes do not know in advance whether it contains at least one element. For instance, we can be interested in the set of real roots of an algebraic

equation, but further investigation can show that it has no real roots at all. In this connection we introduce the concept of an *empty* (or *void* or *null*) set, that is a set containing no elements. The empty set is denoted $\emptyset$, and hence if we write

$$A = \emptyset$$

we mean that A is an empty set. An empty set is considered to be a subset of any set, i.e. relation $\emptyset \subset A$ holds for any set A .

Now we shall give the definitions of some operations which are often performed on sets. More precisely, we shall define not the operations themselves but the sets resulting from these operations.

The *sum* or *union* of two sets A and B is the set consisting of all elements which belong to at least one of the sets A and B ; it is denoted as

$$A \cup B$$

The union of any number of sets is defined analogously. If we are given an arbitrary *indexed family of sets*, that is a collection of sets A_α in which the *index* α runs through an arbitrary set (the *index set*) and does not necessarily take integral values, the union of the sets A_α is denoted

$$\bigcup_{\alpha} A_{\alpha}$$

By definition, this union consists of all elements which belong to at least one of sets A_α . Here and henceforth we follow the convention that if there is no indication of the range of the index α under the sign of union, the union extends over all the values of α which can be taken in the given problem. The same convention refers to the notation of the intersection of sets introduced below.

Let us consider two examples.

1. Let the set A consist of all even numbers, the set B consist of all natural numbers divisible by 3 and the set C of all odd natural numbers not divisible by 3:

$$A = \{2, 4, 6, 8, \dots\}, \quad B = \{3, 6, 9, 12, \dots\}, \\ C = \{1, 5, 7, 11, \dots\}$$

Then their union $A \cup B \cup C$ coincides with the set of all natural numbers.

2. Let, for every real number x , the set A_x consist of all points of the xy -plane having the given abscissa x .^{*} Then the union $\bigcup_x A_x$ is the collection of all the points of the xy -plane.

The *intersection* (or *product* or *meet*) of two sets A and B is the set consisting of all elements which belong both to A and to B . It is denoted by

$$A \cap B$$

We similarly define the intersection

$$\bigcap_{\alpha} A_{\alpha}$$

of any number of sets A_{α} : this is a set of all elements each of which belongs to every A_{α} .

For the sets A and B in Example 1 it follows that their intersection consists of all numbers divisible by 6.

If $B \subset A$ we have

$$A \cup B = A \text{ and } A \cap B = B$$

By the *difference*

$$A \setminus B$$

of two sets A and B is meant the subset of the set A consisting of all the elements of A which do not belong to B . In the definition of the difference $A \setminus B$ it is not necessarily required that B should be a subset of A .

It can be easily verified that

$$A \setminus B = A \setminus (A \cap B)$$

If (and only if) $B \subset A$ we have

$$(A \setminus B) \cup B = A$$

The three notions (i.e. union, intersection and difference of sets) defined above are illustrated in Fig. 1 where by A and B are meant the sets of all points belonging to the corresponding intervals.

Let us prove the so-called *distributive law*: for any family of sets the equality

$$\left(\bigcup_{\alpha} A_{\alpha}\right) \cap B = \bigcup_{\alpha} (A_{\alpha} \cap B) \quad (1)$$

holds. In other words, *intersection distributes over union*.

^{*} Thus, the elements of each set A_x fill a straight line parallel to the y -axis.

We shall establish the desired equality with the help of two opposite inclusion relations. Let $x \in (\bigcup_{\alpha} A_{\alpha}) \cap B$. This means that $x \in \bigcup_{\alpha} A_{\alpha}$ and $x \in B$. Hence, by the definition of union, $x \in A_{\alpha}$ for a certain α ; then $x \in A_{\alpha} \cap B$ and, consequently, $x \in \bigcup_{\alpha} (A_{\alpha} \cap B)$. Thus,

$$(\bigcup_{\alpha} A_{\alpha}) \cap B \subset \bigcup_{\alpha} (A_{\alpha} \cap B)$$

Conversely, let $x \in \bigcup_{\alpha} (A_{\alpha} \cap B)$. This means that $x \in A_{\alpha} \cap B$ for some α . Then $x \in A_{\alpha}$ and, consequently, $x \in$

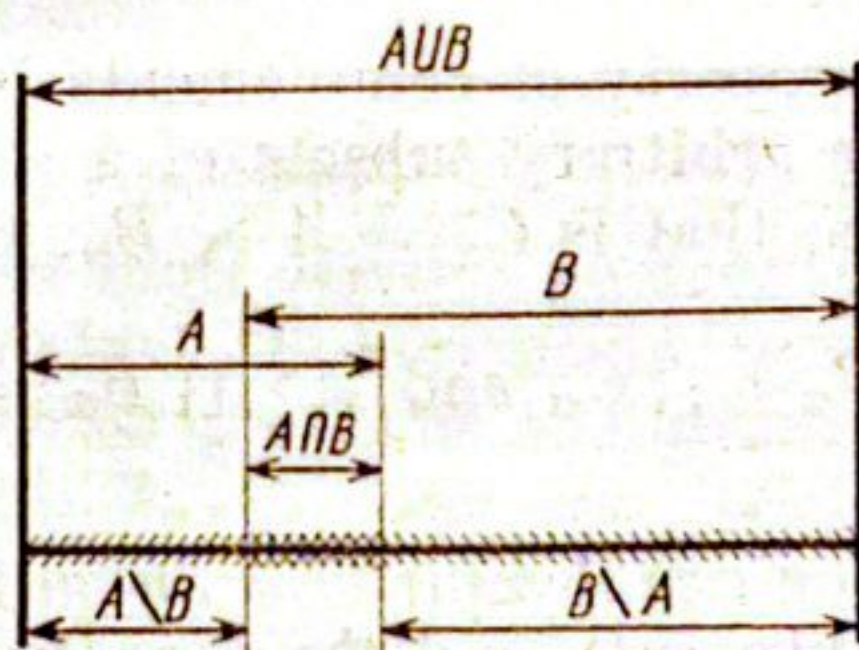


Fig. 1

$\in \bigcup_{\alpha} A_{\alpha}$; besides $x \in B$. Thus, we have $x \in (\bigcup_{\alpha} A_{\alpha}) \cap B$ whence

$$\bigcup_{\alpha} (A_{\alpha} \cap B) \subset (\bigcup_{\alpha} A_{\alpha}) \cap B$$

Equality (1) now follows from the two inclusion relations we have obtained.

We suggest that the reader prove analogously another distributive law

$$(\bigcap_{\alpha} A_{\alpha}) \cup B = \bigcap_{\alpha} (A_{\alpha} \cup B) \quad (1')$$

which means that, conversely, *union distributes over intersection*.

We also have the formula

$$(A \setminus B) \cap C = (A \cap C) \setminus (B \cap C)$$

It should be noted that if $A = \bigcup_{\alpha} A_{\alpha}$ and B is an arbitrary set then

$$A \setminus B = \bigcup_{\alpha} (A_{\alpha} \setminus B)$$

The derivation of the last two formulas is quite simple.

Now we shall define some other notions. Let A be an arbitrary set and $B \subset A$. Then the set $A \setminus B$ is called the *complement* of the set B (with respect to the set A). It is obvious that the complement of $A \setminus B$ is the set B itself. For example, the complement of the set of all rational numbers with respect to the set of all real numbers is the set of all irrational numbers.

The following property of complements is extremely important: if B_{α} are arbitrary subsets of a set A and C_{α} are their complements, that is $C_{\alpha} = A \setminus B_{\alpha}$, then

$$A \setminus \bigcup_{\alpha} B_{\alpha} = \bigcap_{\alpha} C_{\alpha} \text{ and } A \setminus \bigcap_{\alpha} B_{\alpha} = \bigcup_{\alpha} C_{\alpha} \quad (2)$$

In other words, the *complement of a union equals the intersection of the complements*, and the *complement of an intersection equals the union of the complements*. This is the so-called *duality principle*.

We shall prove the first formula; the second is proved in an analogous way. Let $x \in A \setminus \bigcup_{\alpha} B_{\alpha}$. This means that $x \in A$ but $x \notin \bigcup_{\alpha} B_{\alpha}$. Then, by the definition of union, $x \notin B_{\alpha}$ for every α , and consequently $x \in C_{\alpha}$ for all α and thus $x \in \bigcap_{\alpha} C_{\alpha}$. Therefore

$$(A \setminus \bigcup_{\alpha} B_{\alpha}) \subset \bigcap_{\alpha} C_{\alpha}$$

Conversely, let $x \in \bigcap_{\alpha} C_{\alpha}$. This means that $x \in C_{\alpha}$ for all α . Hence, $x \in A$ but $x \notin B_{\alpha}$ for every α . Then $x \notin \bigcup_{\alpha} B_{\alpha}$ and therefore $x \in A \setminus \bigcup_{\alpha} B_{\alpha}$. Thus

$$\bigcap_{\alpha} C_{\alpha} \subset (A \setminus \bigcup_{\alpha} B_{\alpha})$$

From the two relations proved it follows that

$$A \setminus \bigcup_{\alpha} B_{\alpha} = \bigcap_{\alpha} C_{\alpha}$$

Two sets A and B are called *disjoint* (or *nonintersecting*) if their intersection is empty: $A \cap B = \emptyset$. In other words, this means that the sets A and B have no elements in common. For example, the sets of rational and irrational numbers are disjoint. If there is a system of sets A_{α} and any two sets belonging to this system and having different indices are disjoint (i.e. $A_{\alpha_1} \cap A_{\alpha_2} = \emptyset$ for $\alpha_1 \neq \alpha_2$) the sets A_{α} are said to be *pairwise disjoint* (sometimes simply *disjoint*). For instance, the sets A_x in Example 2 considered above are (pairwise) disjoint.

§ 2. Some Auxiliary Relations

In this section we shall derive a number of relations needed for our further aims. The derivation itself can be regarded as a useful exercise on the basic operations on sets.

1°. If A_i ($i = 1, 2, \dots$) are sets forming a decreasing sequence, that is

$$A_1 \supset A_2 \supset \dots \supset A_i \supset \dots$$

and if $\bigcap_{i=1}^{\infty} A_i = \emptyset$ then

$$A_1 = (A_1 \setminus A_2) \cup (A_2 \setminus A_3) \cup \dots = \bigcup_{i=1}^{\infty} (A_i \setminus A_{i+1}) \quad (3)$$

Moreover, it is quite obvious that the sets $A_i \setminus A_{i+1}$ are disjoint.

Proof. It is clear that the right-hand member of (3) is contained in A_1 . Conversely, if $x \in A_1$ we find the greatest number i (let it be $i = n$) for which $x \in A_i$. Then $x \in A_n \setminus A_{n+1}$, which proves that x belongs to the right-hand member.

2°. If A_i are disjoint sets and $B_n = \bigcup_{i=n+1}^{\infty} A_i$ then $\bigcap_{n=1}^{\infty} B_n = \emptyset$. Besides, it is obvious that the sets B_n form a decreasing sequence.

Proof. If $x \in B_n$ for some n then $x \in A_{i_0}$ for some $i_0 > n$. But then $x \notin A_i$ for any $i > i_0$ and therefore $x \notin B_n$ if

$n \geq i_0$. Thus, there are no elements belonging to all B_n simultaneously.

3°. If $A = \bigcup_{i=1}^{\infty} A_i$ then

$$A = A_1 \cup [A_2 \setminus A_1] \cup [A_3 \setminus (A_1 \cup A_2)] \cup \dots$$

$$\dots \cup \left[A_i \setminus \bigcup_{j=1}^{i-1} A_j \right] \cup \dots = \bigcup_{i=1}^{\infty} \left[A_i \setminus \bigcup_{j=1}^{i-1} A_j \right] \quad (4)$$

Furthermore, it is apparent that the sets $A_i \setminus \bigcup_{j=1}^{i-1} A_j$ are disjoint.

Proof. It is only required to prove that the left-hand member of (4) is contained in the right-hand one since the converse is obvious. Let $x \in A$. Then there exists the least number i (let it be i_0) for which $x \in A_i$. If $i_0 = 1$ we have $x \in A_1$. If $i_0 > 1$ then $x \in A_{i_0} \setminus \bigcup_{j=1}^{i_0-1} A_j$. In both cases x belongs to the right-hand member of formula (4).

Formula (4) is notably simplified if the sets A_i form an increasing sequence, that is

$$A_1 \subset A_2 \subset \dots \subset A_i \subset \dots$$

Then $\bigcup_{j=1}^{i-1} A_j = A_{i-1}$, and formula (4) turns into

$$A = A_1 \cup (A_2 \setminus A_1) \cup \dots \cup (A_i \setminus A_{i-1}) \cup \dots \quad (5)$$

§ 3. The Power of a Set

Finite sets can always be compared quantitatively by comparing the numbers of their elements. To this end we can count directly the numbers of the elements or, alternatively, try to set up a *one-to-one correspondence* between the elements of the given sets. For instance, to compare the number of students and the number of seats in a lecture hall it is sufficient to observe whether there are any vacant seats or stu-

* If $i = 1$ the symbol $\bigcup_{j=1}^{i-1}$ denotes the union of an empty family of sets and therefore it is natural to put $\bigcup_{j=1}^{i-1} A_j = \emptyset$ for $i = 1$.

dents with no place to sit down. If all the students can be seated and no empty seats are left there is a one-to-one correspondence between the elements of the set of students and the elements of the set of seats, which means that the number of students coincides with the number of seats. If otherwise, we readily see whether the number of students is greater or smaller than that of the seats. Obviously, the first approach, that is counting elements, is only applicable to finite sets and loses sense for infinite sets. At the same time, the second approach, that is setting up a one-to-one correspondence, can be generalized mathematically to achieve a "quantitative" comparison of infinite sets as well. It then turns out that infinite sets can be "saturated" differently with their elements.

Now we pass to strict definitions. Given two sets A and B , we say that there is a *one-to-one correspondence* between their elements if it is possible to indicate a rule assigning to each element a of A an element b of B (called the *image* of the element a) so that the following two conditions are fulfilled:

- (a) any two different elements of A have different images;
- (b) every element of B is the image of an element of A .*

Definition. Two sets A and B are said to be *equivalent* or to have the *same power* (written $A \sim B$) if it is possible to set up a one-to-one correspondence between their elements.

It should be noted that we do not define the term "power". We only define what is meant by saying that two sets have the same power. Further we shall also define what is meant when one set is said to be of a greater power than the other.**

It is clear that two *finite* sets are equivalent if and only if they include the same number of elements. Let us discuss some examples of equivalent infinite sets.

1. Consider the set N of all natural numbers and the set N_1 of all negative integers. A one-to-one correspondence

* The notion of a one-to-one correspondence plays an important role in the construction of an inverse function. Namely, if $y = f(x)$ is a one-valued function then for its inverse function to be one-valued too it is necessary and sufficient that the correspondence between the values of x and y set up by the function $y = f(x)$ be one-to-one.

** Instead of "power" (of a set) the term "cardinal number" is synonymously used.—Tr.

between their elements is obtained, for example, if the number $-n$ is assigned to every natural number n , and hence these sets are equivalent.

2. Let us take the set N of all natural numbers and the set P of all even positive integers. If to every $n \in N$ we assign the number $2n \in P$ there appears a one-to-one correspondence between the elements of the sets N and P . Thus, $N \sim P$. It should also be noted that $P \subset N$ but $P \neq N$,

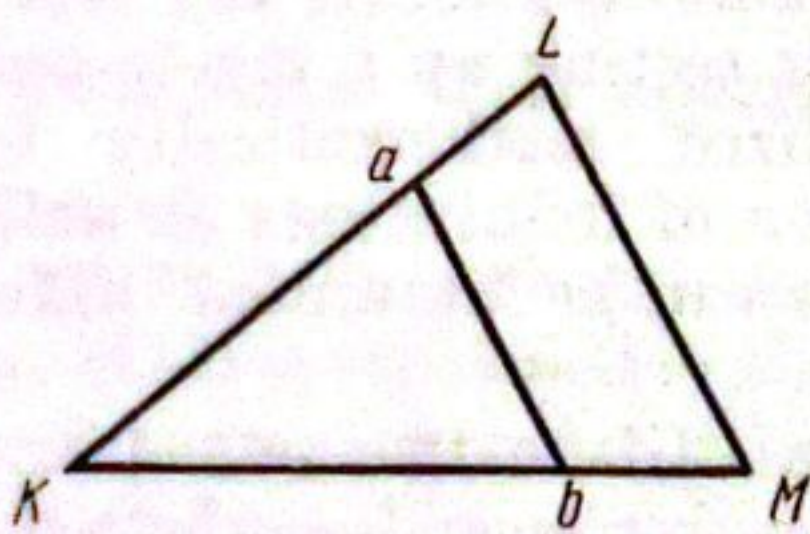


Fig. 2

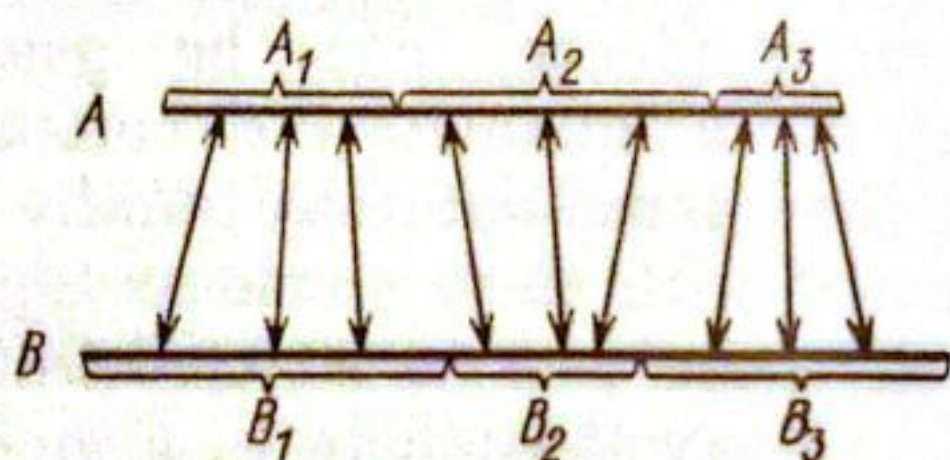


Fig. 3

which shows that an infinite set can be equivalent to its proper subset. Such a situation cannot occur in the case of finite sets.

3. Let E be the set of all real numbers and I be the set of all real numbers belonging to the interval $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. The equivalence $E \sim I$ is shown, for example, with the aid of the correspondence

$$y = \tan x \quad (x \in I, \quad y \in E)$$

4. Let $\triangle KLM$ be a triangle of an arbitrary shape and let A and B be the sets of all the points on its sides KL and KM , respectively. Taking an arbitrary point $a \in A$ we draw through it a straight line parallel to the side LM (Fig. 2) and regard the point of intersection b of this line with KM as the image of the point a . This correspondence between the points of the sides KL and KM is obviously one-to-one and consequently $A \sim B$.

It should be noted that we can take as KL and KM arbitrary line segments of any length. Thus, any two line segments (regarded as point sets) are equivalent irrespective of their lengths.

Let us consider two propositions which will be often used in what follows.

1°. If $A \sim B$ and $B \sim C$ then $A \sim C$ (i.e. two sets which are equivalent to a third set are equivalent to each other).

2°. Let $A = \bigcup_{\alpha} A_{\alpha}$ where the sets A_{α} are pairwise disjoint and let $B = \bigcup_{\alpha} B_{\alpha}$ (the range of α , i.e. the index set, is the same in both cases!) where the sets B_{α} are also disjoint. If $A_{\alpha} \sim B_{\alpha}$ for each α then $A \sim B$.

Proposition 1° is almost obvious. Indeed, suppose that there is a one-to-one correspondence between the elements of the sets A and B and a one-to-one correspondence between the elements of the sets B and C . Let $a \in A$, the image of a being $b \in B$ and the image of the element b being $c \in C$. It can easily be verified that if to each $a \in A$ we assign the element $c \in C$ indicated above we thus set up a one-to-one correspondence between the elements of the sets A and C .

To prove proposition 2°, note that the one-to-one correspondences between the elements of the pairs A_{α}, B_{α} which are set up for all the values of α naturally generate a one-to-one correspondence between the elements of A and B (see Fig. 3 where α assumes the values 1, 2 and 3). Proposition 2° will be referred to as "gluing principle".

§ 4. Countable Sets

In this section we shall consider infinite sets which in some respects are the simplest in comparison with other infinite sets.

Definition. A set is called *countable* if it is equivalent to the set of all natural numbers.*

If N is a set of all natural numbers and A is a countable set, that is $A \sim N$, there exists a one-to-one correspondence between the elements $a \in A$ and the numbers $n \in N$. Thus, to each $a \in A$ there corresponds a number n , and the elements of the set A can be written as $a_1, a_2, \dots, a_n, \dots$. Here a_n denotes the element of A to which the number n corresponds. Hence, the elements of the set A can be arran-

* A countable set is sometimes synonymously termed "a set of power $\aleph_0$ " ($\aleph_0$ is read "aleph-null" or "aleph-zero"). — *Tr.*

ged as an infinite sequence. Conversely, if a set A is such that its elements form an infinite sequence $a_1, a_2, \dots, a_n, \dots$ the numbering of the elements sets up a one-to-one correspondence between the elements of the sets A and N , and hence $A \sim N$. Thus, *countable sets can be characterized as those infinite sets whose elements can be numbered with the aid of all natural numbers.*

It should be noted that the elements of any finite set can also be numbered but in this case not all natural numbers will be used.

Let us consider some examples of countable sets.

1. The set of all even positive integers is countable because even numbers can, for instance, be arranged in an increasing order as

$$a_1 = 2, \quad a_2 = 4, \quad \dots, \quad a_n = 2n, \quad \dots$$

2. The set of all integers is also countable. These numbers cannot be arranged in either increasing or decreasing sequence but they can, for instance, be numbered as follows:

$$a_1 = 0, \quad a_2 = 1, \quad a_3 = -1, \quad a_4 = 2, \quad a_5 = -2 \dots \\ \dots, \quad a_{2n} = n, \quad a_{2n+1} = -n, \quad \dots$$

Let us establish some properties of countable sets.

Theorem I.4.1. *Any infinite set has a countable subset.*

Proof. Let A be an infinite set. We take its arbitrary element and denote it as a_1 . Besides a_1 there is an infinitude of elements of A different from a_1 . Let us choose one of the remaining elements and denote it as a_2 . Next we take an arbitrary element of A different from a_1 and a_2 and denote it by a_3 . Proceeding in this way indefinitely we obtain a countable subset of elements $a_1, a_2, \dots, a_n, \dots$ belonging to A .

Theorem I.4.2. *Every infinite subset of a countable set is countable.*

Proof. Suppose we are given a countable set $A = \{a_1, a_2, \dots, a_n, \dots\}$. Let us take an arbitrary infinite subset $B \subset A$ and arrange all its elements as a sequence with increasing indices:

$$a_{n_1}, a_{n_2}, \dots, a_{n_k}, \dots \quad (n_1 < n_2 < \dots < n_k < \dots)$$

Theorem 1.4.3. *The union of a finite number of countable sets is itself a countable set.*

Let us arrange the elements of the sets A_i as an infinite table of the form*:

Now we can renumber all the elements of Table (6) arranging them, for instance, as the sequence

In other words, we first write down all the elements of the first column, then all the elements of the second column and so on. If the sets A_i contain some common elements one and the same element may occur in sequence (7) several times. However, we naturally number it only once, for instance, when it is first encountered in sequence (7); when this element is encountered repeatedly we simply omit it. Thus, all the elements of the set A can be numbered, that is the set A is countable.

Theorem 1.4.4. *The union of a countable number of countable sets is also a countable set.*

Let us arrange the elements of the sets A_i as a table analogous to Table (6) but containing an infinite number of rows. The elements of this table can also be numbered; for instan-

* In the notation a_{in} of the elements in the table the first index is the number of the set A_i to which this element belongs and the second is the number of the element within the set A_i .

ce, we can count them "diagonally":

$$a_{11}, a_{12}, a_{21}, a_{13}, a_{22}, a_{31}, \dots$$

As in the proof of the previous theorem, repeated elements are numbered only once. Thus, all the elements of the set A can be numbered, that is the set A is countable.

Remark. It is clear that if some of the sets A_i entering into the union are finite, i.e. in the corresponding rows of the table of elements a_{in} only a finite number of places are occupied, this does not prevent us from numbering the elements of the table in the same order as above. Therefore Theorems I.4.3 and I.4.4 continue to hold in the case when some of the sets (but not all) entering into the union are finite. If all of them are finite the union can be finite or countable. A set which is known to be finite or countable is said to be *at most countable*. For brevity, the union of a finite or countable number of sets is called, respectively, finite or countable. Combining what has been said with Theorems I.4.3 and I.4.4 we arrive at the following general result:

Theorem I.4.5. *A finite or countable union of sets each of which is at most countable is itself at most countable.*

Theorem I.4.6. *The set D of all rational numbers is countable.*

Proof. Every nonzero rational number can be written as a fraction m/n in lowest terms where n is a natural number and m is a positive or negative integer*. For a given n , the set of all fractions m/n (m is an integer) is countable (see Example 2). Therefore the set A_n of all fractions m/n in lowest terms is also countable (see Theorem I.4.2).** The set D is the union of all A_n ($n = 1, 2, \dots$) and of the set consisting of one number 0. Since D is an infinite set Theorem I.4.5 indicates that it is countable.

Corollary. *The set of all rational numbers contained in any given interval of the number line is countable.*

It should be noted that rational numbers belonging to an interval cannot of course be arranged as an increasing sequence. For instance, this follows from the fact that there

* An integer $m \neq 0$ is representable as the fraction $m/1$.

** There is obviously an infinitude of fractions in lowest terms with a given denominator.

is no smallest rational number among the numbers exceeding a given rational number. The way of numbering can be found from the proofs of Theorem I.4.6 and of the foregoing theorems on which Theorem I.4.6 is based.

Theorem I.4.7. *If $A = B \cup C$ where B is an arbitrary infinite set and C is at most countable then $A \sim B$ (i.e. the union of an arbitrary infinite set and a finite or a countable set is equivalent to the original infinite set).*

Proof. It can be assumed, without loss of generality, that B and C are disjoint. Let us take a countable subset D of B ; then the sets B and A can be represented as

$$B = (B \setminus D) \cup D \text{ and } A = (B \setminus D) \cup (D \cup C)$$

Thus, the sets A and B are represented as sums of two sets in which the first summands coincide and the second summands are equivalent since D and $D \cup C$ are countable. Therefore, by the gluing principle (see Sec. I.3), $A \sim B$.

Theorem I.4.8. *If A is an uncountable infinite set* and B is its finite or countable subset then $A \setminus B \sim A$.*

Proof. We have $A = (A \setminus B) \cup B$. It is obvious that the set $A \setminus B$ is infinite** and therefore the relation $A \sim (A \setminus B)$ is implied by the previous theorem.

Theorem I.4.9. *Let the elements of a set A be specified by a finite number of parameters each of which can independently take on any value belonging to a countable set. Then the set A is countable.*

Proof. Let us write down the elements of the set A as $a_{p_1, p_2, \dots, p_k}$ where $p_1, p_2, \dots, p_k$ are parameters. We can assume, without loss of generality, that the values of the parameters are natural numbers. For each $a_{p_1, p_2, \dots, p_k} \in A$ we put $n(a) = p_1 + p_2 + \dots + p_k$. It is obvious that $n(a)$ can be any natural number not smaller than k . Given any $n \geq k$, let us denote by A_n the set of all elements of A for which $n(a)$ assumes the given value n .[†] It is clear that every set A_n is finite and $A = \bigcup_{n=k}^{\infty} A_n$; therefore the set A is countable.

* The existence of uncountable sets will be established in the next section.

** If otherwise, the set A would be at most countable, which contradicts the hypothesis.

Corollary. *The set $\mathcal{P}$ of all algebraic polynomials with rational coefficients is countable.*

Proof. The countability of the set $\mathcal{P}_n$ of all polynomials of a fixed degree n with rational coefficients

$$P(x) = c_0x^n + c_1x^{n-1} + \dots + c_n \quad (c_0 \neq 0)$$

is directly implied by the previous theorem (here the role of the parameters is played by the coefficients). But $\mathcal{P} =$

$= \bigcup_{n=0}^{\infty} \mathcal{P}_n$, and hence $\mathcal{P}$ is also countable.

Remark. The countability of the set of all algebraic polynomials with rational coefficients depending on an arbitrary number of variables is proved completely similarly.

A real number is called *algebraic* if it is a root of an algebraic polynomial with integral coefficients. For example, the number $\sqrt{2}$ is algebraic because it is a root of the polynomial $x^2 - 2$. Since each algebraic polynomial can only have a finite number of distinct real roots the countability of the set $\mathcal{P}$ readily implies that the set of all algebraic numbers is countable.

§ 5. The Power of the Continuum

Here we shall not only prove that there exist infinite uncountable sets but also give a simple example of such a set. The term "uncountable" will only be applied to infinite uncountable sets, i.e. finite sets will not be called uncountable. Thus, we can say that in this section we shall discuss the existence of uncountable sets.

Theorem I.5.1. *The set of all real numbers contained in the interval $[0, 1]$ is uncountable.*

Proof. We shall prove this theorem by contradiction. Let us suppose that all the numbers contained in the interval $[0, 1]$ are arranged in a certain way as a sequence $x_1, x_2, \dots, x_n, \dots$. Dividing the interval $[0, 1]$ into three equal parts we choose from the subintervals obtained the one (let us denote it $[a_1, b_1]$) which does not contain x_1 .^{*} Next we divide the interval $[a_1, b_1]$ into three equal parts and choose from them the interval (let it be $[a_2, b_2]$) which does not

^{*} If there appear two subintervals, not one, which do not contain x_1 , we take either of them.

contain x_2 . Proceeding in this way indefinitely we obtain a sequence of intervals $[a_n, b_n]$ such that each of them (beginning with the second) is one third of the preceding interval and $x_n \in [a_n, b_n]$ for every n . By the nested interval theorem, there exists a number c belonging to all the intervals $[a_n, b_n]$. Since $0 \leq c \leq 1$ and we supposed that *all* the numbers belonging to the interval $[0, 1]$ were numbered, there must be $c = x_n$ for some n . On the other hand, by construction, we have $c \notin [a_n, b_n]$ for this n , and thus we arrive at a contradiction. The theorem has been proved.

Definition. A set A is said to have the *power of the continuum* (denoted c^*) if it is equivalent to the set of all real numbers contained in the interval $[0, 1]$.

It should be noted that *any interval* (closed, open or half-open) with end points a and b (for $a \neq b$) has the power c . Indeed, the equivalence of the intervals $[0, 1]$ and $[a, b]$ (as collections of points) was proved in Example 4 in Sec. I.3. A one-to-one correspondence between the points of these intervals can also be set up by means of the formula

$$y = a + (b - a)x$$

Thus, the interval $[a, b]$ is of power c . If one of the end points a, b (or both) is deleted from the interval $[a, b]$ Theorem I.4.8 indicates that we again obtain a set of the power of the continuum. In particular, the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$ is of power c . The set of all real numbers is equivalent to this interval (see Example 3 in Sec. I.3.). Hence, the *set of all real numbers has the power c* . It can be similarly proved that an arbitrary semi-infinite interval is also of power c .

Theorems I.4.6 and I.4.8 directly imply that the *set of all irrational numbers has the power c* . Analogous considerations show that the *set of all nonalgebraic real numbers* (such numbers are called *transcendental*) *is also of power c* . Thus, we have elaborated a simple proof of the existence of transcendental numbers. However, this argument does not enable us to indicate any concrete transcendental number. Nevertheless, some numbers of this kind are known, for instance,

* Or, sometimes, $\aleph$ (read "aleph").—Tr.

π and e ; however, their transcendence is proved by means of extremely sophisticated methods.

Theorems I.4.3 and I.4.4 are analogous to

Theorem I.5.2. *A finite or countable union of sets each of power c is itself of power c .*

Proof. Let

$$A = \bigcup_n A_n \quad (8)$$

where the union of the sets A_n is finite or countable and each of the sets A_n has the power c (the index n takes on natural values). We shall limit ourselves to considering the case when the sets A_n are disjoint. The general case will be discussed in the next section.

For every n we take the interval $[n-1, n)$. Since every interval has the power of the continuum we have $A_n \sim [n-1, n)$. Therefore, by the "gluing principle" (see Sec. I.3),

$$A \sim \bigcup_n [n-1, n)$$

where the union extends over the same values of n as in formula (8). If there is a finite number p of the sets A_n in formula (8) we have $A = \bigcup_{n=1}^p A_n$ and hence $A \sim [0, p)$;

if their number is infinite then $A = \bigcup_{n=1}^{\infty} A_n$, and consequently $A \sim [0, \infty)$. In both cases A is equivalent to an interval and, consequently, is of power c .

For our further aims we need the representation of real numbers as infinite binary fractions. We shall limit ourselves to numbers x belonging to the interval $[0, 1)$: $0 \leq x < 1$. Every such number can be written in the form

$$x = \sum_{n=1}^{\infty} \frac{a_n}{2^n} \quad (9)$$

where each of the digits a_n can only be equal to 0 or 1. We shall impose the requirement that the sum on the right-hand side of (9) should involve an infinite number of zeros; then representation (9) is unique for every number $x \in [0, 1)$.

Let us describe a procedure for determining a_n 's corresponding to any given x . To this end, we first bisect the interval $[0, 1)$. If $0 \leq x < 1/2$ we put $a_1 = 0$; if $1/2 \leq x < 1$ we put $a_1 = 1$. From the two subintervals thus obtained we take the one containing x , bisect it again and determine a_2 following the same principle: if x falls in the left of the two halves we put $a_2 = 0$ and if in the right one we put $a_2 = 1$. All the subsequent digits a_n are found similarly. For instance, if $x = 5/16$ we readily find

$$a_1 = 0, \quad a_2 = 1, \quad a_3 = 0, \quad a_4 = 1, \quad a_n = 0 \text{ for } n \geq 5^*$$

The uniqueness of the representation of any number $x \in [0, 1)$ by formula (9) can be confirmed, for instance, by the following argument. If some $a_n = 0$ then

$$x < \frac{a_1}{2} + \dots + \frac{a_{n-1}}{2^{n-1}} + \sum_{i=n+1}^{\infty} \frac{1}{2^i} = \frac{a_1}{2} + \dots + \frac{a_{n-1}}{2^{n-1}} + \frac{1}{2^n}$$

The sign of strict inequality is guaranteed here since among the digits $a_{n+1}, a_{n+2}, \dots$ there are necessarily some equal to 0. If $a_n = 1$ (for the same n) while the preceding digits $a_1, a_2, \dots, a_{n-1}$ have the same values as before then

$$x \geq \frac{a_1}{2} + \dots + \frac{a_{n-1}}{2^{n-1}} + \frac{1}{2^n}$$

Consequently, the number x cannot be the same in these two cases.

Formula (9) gives the *binary representation* of the number x . Usually, instead of formula (9), we write x as an infinite *binary fraction*:

$$x = 0.a_1a_2 \dots a_n \dots \quad (10)$$

The condition that there must be an infinite number of zeros in (9) (and (10)) excludes the case when (10) is a repeating binary fraction with infinite run of one's**. Conversely, each

* The same number $5/16$ can also be written as $\frac{5}{16} = \frac{1}{4} + \sum_{n=5}^{\infty} \frac{1}{2^n}$

but this representation includes only a finite number of digits a_n equal to 0 and we do not consider it.

** The above example of the number $5/16$ can be used to demonstrate the general fact that every number written as a repeating binary fraction with infinite run of 1's can be represented by another binary fraction without infinite run of 1's.

infinite binary fraction (without infinite run of 1's) of the form $0.a_1a_2\dots a_n\dots$ serves as the representation of a number $x \in [0, 1)$, namely the number x determined by formula (9). We have thus obtained a one-to-one correspondence between the infinite binary fractions (without infinite run of 1's) and the numbers $x \in [0, 1)$. Consequently, the set of all such fractions is of power c .

On the other hand, every infinite binary fraction (10) is completely characterized by the binary places, standing to the right of the binary point, which contain 0. Writing down the numbers of these places as an increasing sequence $n_1 < n_2 < \dots n_k < \dots$ we set up a one-to-one correspondence between the binary fractions (10) (without infinite run of 1's) and all the possible increasing infinite sequences of natural numbers. Consequently, the set H of such sequences is of power c .

Lemma I.5.1. *The set E of all the possible infinite sequences of natural numbers*

$$m_1, m_2, \dots, m_k, \dots^*$$

has the power of the continuum.

Proof. Let $\{m_k\} \in E$. Putting

$$n_1 = m_1, \quad n_2 = m_1 + m_2, \quad \dots$$

$$\dots, \quad n_k = m_1 + m_2 + \dots + m_k, \quad \dots$$

we form the sequence $\{n_k\} \in H$ (where H is the set of all increasing sequences of natural numbers). Any sequence $\{n_k\} \in H$ can be regarded as the image of a sequence $\{m_k\} \in E$, namely of the sequence for which

$$m_1 = n_1, \quad m_k = n_k - n_{k-1} \quad \text{for } k \geq 2$$

Consequently, we have set up a one-to-one correspondence between the sequences belonging to E and those belonging to H . Hence, E and H are of the same power, which is what we wanted to prove.

Theorem I.5.3. *Let the elements of a set A be specified by a finite number of parameters each of which independently assumes any value belonging to a set of power c . Then the set A itself is also of power c .*

* Here we mean arbitrary sequences, not only increasing ones.

in the interval $0 \leq t \leq 1$. This correspondence can be expressed by formulas

$$x = \varphi(t), \quad y = \psi(t)$$

resembling parametric equations of a curve. But it can be proved that if a mapping of the interval $[0, 1]$ onto the unit square is one-to-one, the functions φ and ψ cannot simultaneously be continuous.

Theorem I.5.2 is essentially complemented by the following theorem.

Theorem I.5.4. *If $A = \bigcup_x A_x$ where x runs through a set of power c and each of the sets A_x is also of power c the set A is itself of power c .*

Proof. We can suppose, without losing generality, that x takes all the possible real values. Let us confine ourselves to the case when the sets A_x are disjoint. Each A_x is equivalent to the set of all the points of the xy -plane having the given abscissa x . According to the "gluing principle" (Sec. I.3) the set A is equivalent to the union "of all these straight lines", that is to the set of all points in the plane. Consequently, A is of power c .

§ 6. Comparison of Powers

In the foregoing sections, when proving the equivalence of various sets, we tried to set up a one-to-one correspondence between their elements. But the equivalence of two sets can also be established by means of the following criterion (avoiding the setting up a one-to-one correspondence) which we state without proof: *if a set A is equivalent to a subset $B_1 \subset B$ of a set B while the set B is equivalent to a subset $A_1 \subset A$ then $A \sim B$.* From this criterion it follows immediately that if a subset of a set of power c contains a narrower subset of power c it is itself of power c .*

Using this consequence of the above criterion we can readily extend the proofs of Theorems I.5.2 and I.5.4 to the case when the given sets can intersect.

* For example, every set of points in the plane containing entirely a straight line is of power c .

If we are given two sets A and B and if $A \sim B_1 \subset B$ it may happen that $A \sim B$. If $A \sim B_1 \subset B$ but A is not equivalent to B we say that *the power of the set A is less than that of the set B* (or that *the latter power is greater than the former*).

According to this definition, the power of a countable set is less than the power of any uncountable set and, in particular, is less than the power c . On the other hand, it can be proved that the powers of any two sets are comparable with each other in the sense that either they are equal or one of them is less than the other. Thus, given any set, we can assert that it is of power less than c or is of power c or is of power greater than c . In this connection it appears natural to pose the question ("continuum problem") as to whether there exist uncountable sets of power less than c . G. Cantor suggested the hypothesis ("continuum hypothesis") that there are no such sets, that is c is the least of the powers of uncountable sets. Further development of the set theory showed that the continuum problem itself is connected with some logical difficulties and requires a more rigorous setting closely related to attempts to replace "naive" set theory by a more rigorous *axiomatic* set theory. The framework of this book does not allow us to dwell in more detail on these questions*.

It is easy to construct an example of a set whose power is greater than c . For instance, let Φ be the set of all real functions f defined on the interval $[0, 1]$. The set Φ includes, in particular, all the functions identically equal to a constant which constitute a subset Φ_1 of power c of the set Φ . If we manage to show that the power of the set Φ itself is distinct from c this will exactly mean that it is greater than c .

Let us prove by contradiction that the power of Φ is distinct from c . Suppose that $\Phi \sim [0, 1]$. Then there exists a one-to-one correspondence between the functions f belonging to Φ and the numbers $t \in [0, 1]$. Let us equip each function f with the subscript t equal to the number corresponding to it: f_t . Now we construct the function $g(x) = f_x(x) + 1$ (here

* On modern approaches to continuum problem we refer the reader to the book by P. J. Cohen, *Set Theory and the Continuum Hypothesis*, New York, 1966.

the subscript changes together with the argument). This function belongs to Φ and therefore must coincide with one of the functions f_t . But putting $x = t$ for any t we obtain $g(t) = f_t(t) + 1 \neq f_t(t)$ and thus arrive at a contradiction. This proves that Φ is not equivalent to $[0, 1]$ and, consequently, the power of the set Φ is greater than c .

It turns out that for any set A it is possible to construct a set of greater power. Namely, the set $\mathfrak{A}$ of all the subsets of the set A has a power greater than that of A . In particular, it can be shown that if the set A is countable the corresponding set $\mathfrak{A}$ is of power c and if A is of power c then $\mathfrak{A} \sim \Phi$.

§ 7. Rings, Semirings and Algebras of Sets

In this section we shall consider some new notions related to the set theory which are not connected with comparison of powers of sets.

Definition. Let M be an arbitrary set. A nonempty system $\mathfrak{A}$ of some subsets of M is called a *ring (of sets)* if for any sets $A, B \in \mathfrak{A}$ the following two conditions hold:

- (i) $A \cup B \in \mathfrak{A}$;
- (ii) $A \setminus B \in \mathfrak{A}$.

It is clear that condition (i) immediately extends, by induction, to any finite number of sets contained in $\mathfrak{A}$, that is if $A = \bigcup_{i=1}^n A_i$ and all $A_i \in \mathfrak{A}$ then also $A \in \mathfrak{A}$.

Thus, a ring can be characterized as a nonempty system of some subsets, of a given set, closed with respect to the operations of forming unions of finite number of sets and differences of sets.

Trivial examples of a ring are the collection of all subsets of a given set M and the system consisting of only one empty subset.* The family of all finite subsets of M (the empty set is also regarded as a finite subset) is also a ring. Later on we shall encounter some more significant examples of rings.

Any ring of sets $\mathfrak{A}$ contains an empty set. Indeed, if $A \in \mathfrak{A}$ (such a set A must exist since $\mathfrak{A}$ is nonempty) then $A \setminus A = \emptyset$ and hence $\emptyset \in \mathfrak{A}$.

* The latter system is nonempty since it contains one element $\emptyset$.

Any ring $\mathfrak{A}$ is closed with respect to the operation of forming intersections of finite number of sets. It suffices to verify this fact for the intersection of two arbitrary sets. Let $A, B \in \mathfrak{A}$. Then from the formula

$$A \cap B = A \setminus (A \setminus B)$$

and from the definition of a ring it follows directly that $A \cap B \in \mathfrak{A}$.

Definition. A nonempty system $\mathfrak{A}$ of subsets of a set M is called an *algebra* if

- (i) $A \cup B \in \mathfrak{A}$ whenever $A, B \in \mathfrak{A}$,
- (ii) for any $A \in \mathfrak{A}$ its complement $C = M \setminus A$ also belongs to $\mathfrak{A}$.

Theorem I.7.1. A system $\mathfrak{A}$ of subsets of a set M is an algebra if and only if it is a ring and $M \in \mathfrak{A}$.

Proof. Let $\mathfrak{A}$ be an algebra. Choosing any $A \in \mathfrak{A}$ we can write

$$M = A \cup (M \setminus A)$$

and therefore $M \in \mathfrak{A}$. It only remains to show that $\mathfrak{A}$ is closed with respect to the operation of forming differences of sets. Let $A, B \in \mathfrak{A}$. We have

$$A \setminus B = A \cap (M \setminus B) = M \setminus [(M \setminus A) \cup B]^*$$

and, since $M \setminus A, B \in \mathfrak{A}$, the set $A \setminus B$ also belongs to $\mathfrak{A}$.

Conversely, if $\mathfrak{A}$ is a ring and $M \in \mathfrak{A}$ the complement $M \setminus A$ of any set $A \in \mathfrak{A}$ also belongs to $\mathfrak{A}$ and thus $\mathfrak{A}$ is an algebra of sets.

Remark. In condition (i) in the definition of an algebra the union of sets can be replaced by the intersection.** Indeed, if $A \cap B \in \mathfrak{A}$ whenever $A, B \in \mathfrak{A}$ then we also have

$$A \cup B = M \setminus [(M \setminus A) \cap (M \setminus B)] \in \mathfrak{A}$$

Definition. A nonempty system $\mathfrak{A}$ of subsets of a set M is termed a σ -ring if it is a ring closed with respect to the operation of forming not only finite but also countable unions of sets, that is

* This equality follows from each of the formulas (2).

** In the definition of a ring such a replacement is illegitimate.

(i) $A_i \in \mathfrak{A}$ ($i = 1, 2, \dots$) implies

$$A = \bigcup_{i=1}^{\infty} A_i \in \mathfrak{A}^*$$

(ii) $A, B \in \mathfrak{A}$ implies $A \setminus B \in \mathfrak{A}$.

Every σ -ring is also closed with respect to the operation of forming intersections of a countable number of sets.

Indeed, if $A_i \in \mathfrak{A}$ ($i = 1, 2, \dots$) and $A = \bigcap_{i=1}^{\infty} A_i$ the equality

$$A = \bigcap_{i=1}^{\infty} (A_1 \cap A_i) = A_1 \setminus \bigcup_{i=1}^{\infty} (A_1 \setminus A_i)$$

shows that $A \in \mathfrak{A}$.

The notion of a σ -algebra is defined in like manner: a nonempty system $\mathfrak{A}$ of subsets of M is called a σ -algebra if it satisfies condition (i) entering into the definition of a σ -ring and condition (ii) in the definition of an algebra. Repeating literally the proof of Theorem I.7.1 we can show that for a system $\mathfrak{A}$ to be a σ -algebra it is necessary and sufficient that $\mathfrak{A}$ be a σ -ring and that $M \in \mathfrak{A}$.

The collection of all the subsets of a set M is a trivial example of a σ -algebra. The collection of all countable subsets of the set M and also the system consisting of a single set $\emptyset$ are examples of a σ -ring.

If we are given a family of rings (algebras) $\mathfrak{A}_\alpha$ of some subsets of a set M their intersection $\mathfrak{A} = \bigcap_{\alpha} \mathfrak{A}_\alpha$ is also a ring (an algebra)**. For instance, let us verify that $\mathfrak{A}$ is closed with respect to the addition of sets. If $A, B \in \mathfrak{A}$ then $A, B \in \mathfrak{A}_\alpha$ for any α and, consequently, $A \cup B \in \mathfrak{A}_\alpha$ for any α ; therefore $A \cup B \in \mathfrak{A}$.

Similarly, if $\mathfrak{A}_\alpha$'s are σ -rings (σ -algebras) their intersection is also a σ -ring (a σ -algebra).

* In particular, this condition implies that the union of any finite number of sets belonging to $\mathfrak{A}$ is also contained in $\mathfrak{A}$ because we can take in condition (i) all A_i 's, beginning with i , equal to each other.

** Such an intersection is nonempty since it necessarily contains the empty set $\emptyset \in \mathfrak{A}$.

If $\mathfrak{B}$ is an arbitrary nonempty system of subsets of a set M there always exists the smallest ring (algebra, σ -ring, σ -algebra) containing $\mathfrak{B}$. Indeed, such a ring (algebra, σ -ring, σ -algebra) $\mathfrak{A}$ is the intersection of all the rings (algebras, σ -rings, σ -algebras) $\mathfrak{A}'$ consisting of subsets of the set M and containing $\mathfrak{B}$.^{*} This system of sets $\mathfrak{A}$ is called the *ring (algebra, σ -ring, σ -algebra) generated by the system of subsets $\mathfrak{B}$* .

A more general notion than that of a ring is introduced in the following definition.

Definition. A system $\mathfrak{R}$ of subsets of a set M is called a *semiring* if it satisfies the following three conditions:

- (i) $\emptyset \in \mathfrak{R}$;
- (ii) $A \cap B \in \mathfrak{R}$ whenever $A, B \in \mathfrak{R}$;
- (iii) if $A, B \in \mathfrak{R}$ and $B \subset A$ there exists a finite or countable family of disjoint sets $C_n \in \mathfrak{R}$ such that $A \setminus B = \bigcup_n C_n$.

A simple example of a semiring is the collection of all the intervals in the real line (including the "degenerated" ones, i.e. consisting of a single point; the empty set is also regarded as a "degenerated" interval). The difference of two intervals is either an interval or a union of two disjoint intervals. A more interesting example of a semiring will be treated in Chapter V.

Let us enumerate some properties of semirings.

(a) If $A, A_1, \dots, A_p \in \mathfrak{R}$ where $\mathfrak{R}$ is a semiring there is an at most countable system of disjoint sets $C_n \in \mathfrak{R}$ such that

$$A \setminus \bigcup_{i=1}^p A_i = \bigcup_n C_n$$

This assertion is obviously a strengthened variant of condition (iii) entering into the definition of a semiring.

We shall prove this property by induction. We have

$$A \setminus A_1 = A \setminus (A \cap A_1)$$

and, since $A \cap A_1 \in \mathfrak{R}$, we can write $A \setminus A_1 = \bigcup_k D_k$ where $D_k \in \mathfrak{R}$ are disjoint sets and the range of k is at most count-

^{*} Such $\mathfrak{A}'$ always exist; for instance, as $\mathfrak{A}'$ we can take the collection of all the subsets of the set M .

able. Furthermore,

$$A \setminus (A_1 \cup A_2) = (A \setminus A_1) \setminus A_2 = \bigcup_h (D_h \setminus A_2)$$

By what has already been proved, each of the sets $D_h \setminus A_2$ is representable as a finite or countable union of disjoint sets $E_{hl} \in \mathfrak{R}$:

$$D_h \setminus A_2 = \bigcup_l E_{hl}$$

Therefore, since D_h 's are disjoint, all the sets E_{hl} are also disjoint and

$$A \setminus (A_1 \cup A_2) = \bigcup_{h,l} E_{hl}$$

Now it becomes clear that this argument can be repeated any required number of times.

(b) If a set C is an at most countable union of sets $A_n \in \mathfrak{R}$, i.e. $C = \bigcup_n A_n$, the set C can be represented in the form of an at most countable union of disjoint sets $B_m \in \mathfrak{R}$, i.e. $C = \bigcup_m B_m$, such that each B_m is contained in at least one of the sets A_n .

To prove the latter property we make use of 3°, Sec. I.2* and represent C as a union of disjoint sets:

$$C = \bigcup_n \left(A_n \setminus \bigcup_{j=1}^{n-1} A_j \right)$$

Next, each of the sets $A_n \setminus \bigcup_{j=1}^{n-1} A_j$ is replaced by a union of disjoint sets $D_{nh} \in \mathfrak{R}$ (according to proposition (a), this is possible). The system of all D_{nh} 's is at most countable. Lastly, renumbering the sets D_{nh} as a single sequence we obtain the desired sets B_m .

* Proposition 3° in Sec. I.2 obviously applies not only to countable but to finite unions as well.

Point Sets in a Euclidean Space

§ 1. n -dimensional Euclidean Space

As is known, the introduction of a coordinate system makes it possible to identify the points in the plane with the ordered pairs of real numbers and the points of the three-dimensional (geometric) space with the triples of real numbers. In this interpretation the three-dimensional space can be regarded as the set of all ordered triples (x, y, z) of real numbers. Further generalization of this approach leads to the notion of a space of an arbitrary finite dimension n whose "points" are the ordered n -tuples of real numbers; such spaces prove useful for studying functions of several variables. This chapter will be devoted to the investigation of such spaces.

Definition. Let n be an arbitrary natural number. By a *point of the n -dimensional space* (for brevity, we shall simply speak of a *point*) we mean an ordered collection of n real numbers $\xi_1, \xi_2, \dots, \xi_n$, that is an n -tuple $(\xi_1, \xi_2, \dots, \xi_n)$. Such a point will be denoted by a single letter, for instance as

$$x = (\xi_1, \xi_2, \dots, \xi_n)$$

The numbers $\xi_1, \xi_2, \dots, \xi_n$ are called the *coordinates* of the point x . The set of all points $x = (\xi_1, \xi_2, \dots, \xi_n)$ (for a given n) is called the *n -dimensional (Cartesian) space* (or simply, *n -space*).

If $n = 1$ every point is specified by exactly one coordinate. Therefore the one-dimensional space can be identified with the set of all real numbers or with the set of all the points on the line.

From Theorem 1.5.3 we conclude that the set of all the points of n -space (for any n) is of power c . On the other hand, the set of all rational points of an n -space (by which

we mean the points whose all coordinates are rational numbers) is countable (see Theorem I.4.9).

In the n -dimensional space of n -tuples we can introduce in a natural way the notion of the *origin* (as the point with all coordinates equal to zero: $\xi_1 = \xi_2 = \dots = \xi_n = 0$) and the notion of *coordinate axes*. For instance, by the first coordinate axis (the ξ_1 -axis) we mean the subset consisting of all points for which ξ_1 can assume arbitrary values while $\xi_2 = \dots = \xi_n = 0$. Furthermore, in calculus and in the theory of functions we need the notion of a distance between the points of the n -space. To introduce this important notion we generalize the well-known formula of analytic geometry and define the *distance* between two points $x = (\xi_1, \xi_2, \dots, \xi_n)$ and $y = (\eta_1, \eta_2, \dots, \eta_n)$ as the nonnegative number (denoted $\rho(x, y)$) specified by the formula

$$\rho(x, y) = \sqrt{\sum_{i=1}^n (\xi_i - \eta_i)^2} \quad (1)$$

There are also some other ways of introducing the distance in an n -dimensional space. In our course we shall use formula (1) which is a direct generalization, to the n -dimensional case, of the distance between the points of the three-dimensional space studied in the Euclidean geometry.

Definition. The n -dimensional space of ordered n -tuples with the distance between its points determined by formula (1) is called the *n -dimensional Euclidean space* (or, simply, *Euclidean n -space*) and is denoted by R_n .

It is clear that $\rho(x, y) = 0$ if and only if $x = y$, i.e. if and only if $\xi_i = \eta_i$ for all $i = 1, 2, \dots, n$. It is also obvious that $\rho(y, x) = \rho(x, y)$. Let us prove that for any three points $x, y, z \in R_n$ the inequality

$$\rho(x, y) \leq \rho(x, z) + \rho(z, y) \quad (2)$$

holds. In the cases of dimension 2 or 3, that is for the points in the plane or in the three-dimensional geometric space, inequality (2) expresses the well-known fact that the sum of the lengths of any two sides of a triangle is not less than the length of the third side.* This accounts for the term "the

* The sign of equality only appears when the triangle "degenerates", that is when all its three vertices lie in one line.

triangle inequality" applied to formula (2) in the general case of an arbitrary dimension n .

Before proving (2) we shall establish an important relation known as *Cauchy's* inequality*:

$$\left(\sum_{i=1}^n a_i b_i\right)^2 \leq \left(\sum_{i=1}^n a_i^2\right) \left(\sum_{i=1}^n b_i^2\right) \quad (3)$$

(here a_i and b_i ($i = 1, 2, \dots, n$) are arbitrary real numbers). A simple proof of this inequality is based on the following property of a quadratic trinomial $Ax^2 + 2Bx + C$ with real coefficients: if it is nonnegative for all real x its discriminant $B^2 - AC$ must also be nonnegative: i.e. $B^2 - AC \leq 0$.** Let us take the auxiliary function $\varphi(x)$ of the real variable x specified by the formula

$$\varphi(x) = \sum_{i=1}^n (a_i x + b_i)^2 = Ax^2 + 2Bx + C$$

where the coefficients of the resulting quadratic trinomial $Ax^2 + Bx + C$ are

$$A = \sum_{i=1}^n a_i^2, \quad B = \sum_{i=1}^n a_i b_i \quad \text{and} \quad C = \sum_{i=1}^n b_i^2$$

The definition of φ shows directly that $\varphi(x) \geq 0$ for all x . Therefore, using the above criterion we can write

$$\left(\sum_{i=1}^n a_i b_i\right)^2 - \left(\sum_{i=1}^n a_i^2\right) \left(\sum_{i=1}^n b_i^2\right) \leq 0$$

(which is nothing but Cauchy's inequality with all terms transposed to the left-hand side).

From inequality (3) we derive another important inequality

$$\sqrt{\sum_{i=1}^n (a_i + b_i)^2} \leq \sqrt{\sum_{i=1}^n a_i^2} + \sqrt{\sum_{i=1}^n b_i^2} \quad (4)$$

* A. L. Cauchy (1789-1857), a prominent French mathematician.

** If $A \neq 0$ and $Ax^2 + 2Bx + C \geq 0$ for all x the trinomial cannot have two distinct real roots and therefore $B^2 - AC \leq 0$. If $A = 0$ the trinomial degenerates into the linear function $2Bx + C$, and if this function does not change sign there must also be $B = 0$, and thus $B^2 - AC = 0$.

(where a_i and b_i are any real numbers) which is also called Cauchy's inequality.

Now, to prove inequality (4) we extract the square roots of both members of inequality (3)*, duplicate both members of the resultant inequality and then add to them the expression $\sum_{i=1}^n a_i^2 + \sum_{i=1}^n b_i^2$. This yields

$$\begin{aligned} \sum_{i=1}^n a_i^2 + 2 \sum_{i=1}^n a_i b_i + \sum_{i=1}^n b_i^2 &\leq \\ &\leq \sum_{i=1}^n a_i^2 + 2 \sqrt{\sum_{i=1}^n a_i^2} \sqrt{\sum_{i=1}^n b_i^2} + \sum_{i=1}^n b_i^2 \end{aligned}$$

On rewriting the last inequality as

$$\sum_{i=1}^n (a_i + b_i)^2 \leq \left(\sqrt{\sum_{i=1}^n a_i^2} + \sqrt{\sum_{i=1}^n b_i^2} \right)^2$$

and extracting the square roots of both members of the latter inequality we arrive at (4).

Now we can readily prove triangle inequality (2). Let

$$\begin{aligned} x &= (\xi_1, \xi_2, \dots, \xi_n), & y &= (\eta_1, \eta_2, \dots, \eta_n), \\ z &= (\zeta_1, \zeta_2, \dots, \zeta_n) \end{aligned}$$

Setting

$$a_i = \xi_i - \zeta_i \text{ and } b_i = \zeta_i - \eta_i \quad (i = 1, 2, \dots, n)$$

in inequality (4) we just obtain inequality (2).

§ 2. Limit Points

Generalizing the notion of the *limit* of a number sequence we state the following

Definition. A point $x \in R_n$ is called the *limit* of a sequence of points $x_m \in R_n$ (as usual, we write $x_m \rightarrow x$ or $x = \lim x_m$) if $\rho(x_m, x) \rightarrow 0$. Let $x_m = (\xi_1^{(m)}, \xi_2^{(m)}, \dots, \xi_n^{(m)})$ and $x =$

* In doing so we take the value $\sum_{i=1}^n a_i b_i$ of the root of the left-hand member which may not coincide with the arithmetic root.

$= (\xi_1, \xi_2, \dots, \xi_n)$. Then from the formula

$$\rho(x_m, x) = \sqrt{\sum_{i=1}^n (\xi_i^{(m)} - \xi_i)^2}$$

it is clearly seen that the relation $x_m \rightarrow x$ is equivalent to the simultaneous relations

$$\xi_i^{(m)} \rightarrow \xi_i \text{ for all } i = 1, 2, \dots, n$$

Thus, the convergence of a sequence of points in R_n is equivalent to the convergence with respect to the coordinates ("coordinate-wise" convergence).

Definition. The open ball (sphere) $S(x_0, \varepsilon)$ with centre at $x_0 \in R_n$ and radius $\varepsilon > 0$ is the set of all points $x \in R_n$ for which $\rho(x, x_0) < \varepsilon$.^{*} Any open sphere with centre at x_0 is called a *neighbourhood* of the point x_0 . A neighbourhood with a radius ε is also termed an ε -neighbourhood.

Note that in the space R_1 the open sphere $S(x_0, \varepsilon)$ turns into the open interval $(x_0 - \varepsilon, x_0 + \varepsilon)$ and in R_2 into the (open) circle (more precisely, disc) $(x - x_0)^2 + (y - y_0)^2 < \varepsilon^2$.

It is easily seen that for the relation $x_m \rightarrow x_0$ to hold it is necessary and sufficient that, given any $\varepsilon > 0$, there should exist M such that $x_m \in S(x_0, \varepsilon)$ for all $m \geq M$.

Now let us take an arbitrary set of points A belonging to R_n . A point $x \in R_n$ is called a *limit point* (or *accumulation point*) of the set A if there exists a sequence of points $x_m \in A$ different from x ($x_m \neq x$) such that $x_m \rightarrow x$. It is clear that such a sequence must contain an infinitude of distinct points^{**}.

For example, if A is the set of all rational points of R_n any point belonging to R_n is a limit point of A . In other

^{*} The confusion between "sphere" understood in the sense of spherical surface and "sphere" meant in the sense of a solid sphere (or ball) can be avoided by use of the adjective "open" (the same refers to the term "closed sphere (ball)" introduced later). By the way, the meaning of the terms will always be clear from the context and we shall not stipulate this distinction.—Tr.

^{**} For, if otherwise, some point x_m would repeat infinitely many times, that is $x_{m_1} = x_{m_2} = \dots = x_{m_h} = \dots$ for $m_1 < m_2 < \dots < m_h < \dots$. But, since $x_m \rightarrow x$, the subsequence x_{m_h} also converges to x whence $x_{m_1} = x_{m_2} = \dots = x_{m_h} = \dots = x$, which contradicts the definition of limit point.

words, any point of R_n is representable as the limit of a sequence of some rational points.

The limit points of a nonempty finite interval $(a, b) \subset R_1$ are all the points of the interval and also its end points $x = a$ and $x = b$.

It should also be mentioned that if a point x belongs to a set A (here A is again an arbitrary subset of R_n) but is not its limit point the point x is called an *isolated point* of the set A .

Limit points can be given another characteristic expressed by the following criterion.

Theorem II.2.1. *For a point $x \in R_n$ to be a limit point of a set $A \subset R_n$ it is necessary and sufficient that every neighbourhood of the point x contain at least one point $y \in A$ distinct from x .*

The proof below incidentally shows that if x is a limit point of a set A its every neighbourhood contains *infinitely many* points of the set A .

Proof. (a) *Necessity.* Let x be a limit point of the set A and S be its arbitrary neighbourhood. By the definition of a limit point, there is a sequence $x_m \rightarrow x$ with all $x_m \neq x$ and all $x_m \in A$. Therefore, beginning with a certain number m , all x_m are contained in S , and there are infinitely many different points among them.

(b) *Sufficiency.* Let the condition of the theorem be fulfilled. Then for any natural number m there is a point $y_m \in A$ distinct from x and belonging to the neighbourhood $S(x, 1/m)$. Therefore $\rho(y_m, x) < 1/m$ and thus $y_m \rightarrow x$. Consequently, x is a limit point of the set A .

Definition. A set $A \subset R_n$ is said to be *bounded* if the coordinates of all $x \in A$ are uniformly bounded (i.e. their moduli are bounded by one and the same positive constant). This is equivalent to the requirement that A be contained within a sphere of a sufficiently large radius.*

* If $|\xi_i| < K$ ($i = 1, 2, \dots, n$) then $x \in S(O, K\sqrt{n})$ where O is the origin. Conversely, if $x \in S(x_0, r)$ then $|\xi_i - \xi_i^{(0)}| < r$ for all i , that is

$$\xi_i^{(0)} - r < \xi_i < \xi_i^{(0)} + r \quad (i = 1, 2, \dots, n)$$

(here $x = (\xi_1, \xi_2, \dots, \xi_n)$ and $x_0 = (\xi_1^{(0)}, \xi_2^{(0)}, \dots, \xi_n^{(0)})$).

As is known from calculus, the *Bolzano-Weierstrass theorem* asserts that any bounded number sequence contains a convergent subsequence having a finite limit.

This theorem extends to sequences of points in R_n as well: any bounded sequence of points $x_m \in R_n$ contains a subsequence convergent to a limit.*

We shall present the proof of the latter proposition. Let $x_m = (\xi_1^{(m)}, \xi_2^{(m)}, \dots, \xi_n^{(m)})$ be some points of R_n forming a bounded sequence. Applying the Bolzano-Weierstrass theorem (which is known to hold for number sequences) to the sequence $\{\xi_1^{(m)}\}$ of the first coordinates we can find a subsequence of the values of the index m , say m_k , for which $\{\xi_1^{(m_k)}\}$ possesses a finite limit: $\xi_1^{(m_k)} \rightarrow \xi_1$. Now, considering only those points of the original sequence $\{x_m\}$ which are numbered by m_k we obtain a subsequence, of the original sequence, in which the number sequence of the first coordinates of its members is convergent. Next, taking the number sequence of the second coordinates of the points x_{m_k} of the latter sequence, that is the number sequence $\{\xi_2^{(m_k)}\}$, we can find a subsequence of the subsequence $\{x_{m_k}\}$ for which the number sequence of the second coordinates of its points is convergent to a finite limit. In this procedure the convergence of the sequence of first coordinates is not violated.

Applying this procedure repeatedly n times we arrive at a subsequence of the given sequence $\{x_m\}$ for which all the number sequences of the corresponding coordinates of its points are convergent to some finite limits. Since the convergence in R_n is equivalent to the coordinate-wise convergence the resulting subsequence of points has a limit in R_n .

Let us establish a theorem whose formulation is in fact a restatement of the above-proved proposition; this theorem is also referred to as the *Bolzano-Weierstrass theorem*.

Theorem II.2.2. *Every bounded infinite set $A \subset R_n$ has at least one limit point. (Of course, this limit point may not belong to A .)*

Proof. Let us choose a countable subset of points x_m of the set A (different from each other). These points form a bounded sequence since the whole set A is bounded. By

* Here we do not stress that the limit is finite since we do not consider other ("infinite") limits in R_n .

the former statement of the Bolzano-Weierstrass theorem, there exists a subsequence $\{x_{m_k}\}$ of the sequence $\{x_m\}$ having a limit: $x_{m_k} \rightarrow x$. Since all x_{m_k} are distinct from each other there are infinitely many points among them different from x , which shows that x is a limit point of the set A .

§ 3. Closed and Open Sets

Definition. A set $F \subset R_n$ is said to be *closed* if, given any convergent sequence of points $x_m \in F$, the limit of the sequence also belongs to F .

This definition implies that the whole space R_n is closed; the empty set is also considered (formally) closed since it contains no sequences of points at all. Any finite set $F = \{y_1, y_2, \dots, y_k\}$ * is closed since if an infinite sequence is formed of points belonging to this set at least one of the points, say y_i , contained among x_m 's repeats infinitely many times, and therefore the limit (provided it exists) must coincide with y_i . **

Examples of closed sets in R_1 are a closed interval $[a, b]$, the set of all numbers $x \geq a$ or the set of all numbers $x \leq b$.

In the space R_2 the following sets are examples of a closed set: any rectangle specified by inequalities $a \leq \xi_1 \leq b$, $c \leq \xi_2 \leq d$; the collection of all the points of a curve $\xi_1 = \varphi(t)$, $\xi_2 = \psi(t)$ ($\alpha \leq t \leq \beta$) where the functions φ and ψ are continuous.

On the other hand, in the space R_n , for any n , the set of all rational points is not closed since the limit of a sequence of rational points can be any point of R_n .

It is clear that *a closed set contains all its limit points*. But it should be noted that in the definition of a closed set we meant arbitrary convergent sequences whose members, in particular, can coincide with the limits of the sequences. It appears obvious that we can state an equivalent definition by calling a set closed when it contains all its limit points.

* This notation indicates that F consists of k points: $y_1, y_2, \dots, y_k$.

** It can easily be understood that in the case of a convergent sequence all the points x_m , beginning with some m , must coincide with y_l .

Let us proceed to establish the basic properties of closed sets. In our considerations below we deal with sets contained in one and the same fixed Euclidean space R_n .

Theorem II.3.1. *The union of a finite number of closed sets is also a closed set.*

Proof. Let F_i ($i = 1, 2, \dots, p$) be closed sets, $F = \bigcup_{i=1}^p F_i$, $x_m \in F$ ($m = 1, 2, \dots$) and $x_m \rightarrow x$. The sequence $\{x_m\}$ contains an infinite subsequence $\{x_{m_k}\}$ composed completely of the points of one of the constituent sets, for instance, $x_{m_k} \in F_j$. But $x_{m_k} \rightarrow x$ and, since F_j is closed, we also have $x \in F_j$ and therefore $x \in F$.

Theorem II.3.2. *The intersection of any family of closed sets is closed.*

Proof. Let $F = \bigcap_{\alpha} F_{\alpha}$ where all F_{α} 's are closed. If $x_m \rightarrow x$ and $x_m \in F$ ($m = 1, 2, \dots$) then $x_m \in F_{\alpha}$ for any α , and hence we also have $x \in F_{\alpha}$ for any α . Consequently $x \in F$, and F is closed.

It should be stressed that, in contrast to Theorem II.3.2, Theorem II.3.1 cannot be generalized to an arbitrary system of closed sets. For instance, the half-open interval $(0, 1]$ belonging to R_1 can be represented as the countable union

$$(0, 1] = \bigcup_{n=2}^{\infty} [1/n, 1]$$

of the closed sets $\left[\frac{1}{n}, 1\right]$ while the interval $(0, 1]$ itself is not closed.

Definition. Let A be an arbitrary set of points belonging to R_n . A point $x \in A$ is called an *interior point* of the set A if there is some $\varepsilon > 0$ such that the neighbourhood $S(x, \varepsilon)$ of the point x is entirely contained in A : $S(x, \varepsilon) \subset A$. A set $G \subset R_n$ is said to be *open* if all its points are interior.

The entire space R_n is an open set; the empty set is also included (formally) in the class of open sets. Examples of an open set in R_1 are any interval of the type (a, b) and the set determined by inequality $x > a$ or $x < b$. An n -dimensional rectangular parallelepiped not including its boundary

which, by definition, is a set specified by inequalities

$$a < \xi_1 < b, \quad c < \xi_2 < d$$

is an example of an open set in $\mathbf{R}_n$.

Let us show that every open sphere $S(x, r)$ in any space $\mathbf{R}_n$ is an open set. Let $y \in S(x, r)$. This means that $\rho(y, x) < r$. Let us put $\varepsilon = r - \rho(y, x)$. If $z \in S(y, \varepsilon)$ we have

$$\rho(z, x) \leq \rho(z, y) + \rho(y, x) < \varepsilon + \rho(y, x) = r$$

whence $z \in S(x, r)$. Thus, the whole open sphere $S(y, \varepsilon)$ lies within $S(x, r)$ and, consequently, y is an interior point of the sphere $S(x, r)$.

The following important theorem describes a relationship between open and closed sets.

Theorem II.3.3. *A set $G \subset \mathbf{R}_n$ is open if and only if its complement $F = \mathbf{R}_n \setminus G$ is closed.*

Proof. (a) *Necessity.* Let G be open, $x_m \in F$ ($m = 1, 2, \dots$) and $x_m \rightarrow x$. If x belonged to G then an entire neighbourhood $S(x, \varepsilon)$ of the point x would also be contained in G and hence would not contain any point of F , which is impossible since x is the limit of the points $x_m \in F$. Consequently $x \in F$, and thus F is closed.

(b) *Sufficiency.* Let F be closed and $x \in G$. If x were not an interior point of the set G its every neighbourhood would contain a point $y \notin G$, that is a point y belonging to F and distinct from x (since $x \notin F$). Therefore, by Theorem II.2.1, x would be a limit point of the set F (since F is closed) and hence the inclusion relation $x \in F$ would hold. We thus arrive at a contradiction proving that x is an interior point of G , and thus G is an open set.

Theorem II.3.4. *The union of any system of open sets is again an open set.*

Proof. Let $G = \bigcup_{\alpha} G_{\alpha}$ where all G_{α} 's are open. Then the sets $F_{\alpha} = \mathbf{R}_n \setminus G_{\alpha}$ are closed and, by the property of complements (see Sec. I.1), $\mathbf{R}_n \setminus G = \bigcap_{\alpha} F_{\alpha}$. Therefore Theorem II.3.2 shows that $\mathbf{R}_n \setminus G$ is closed and, by Theorem II.3.3, G is open.

By a complete analogy with this theorem the following proposition is proved (using Theorem II.3.1).

Theorem II.3.5. *The intersection of a finite number of open sets is also an open set.*

§ 4. The Structure of an Open Set on the Line

In this section we shall deal with point sets on the line, i.e. belonging to R_1 . It follows from Theorem II.3.4 that the union of any system of open intervals, i.e. of the type (a, b) (including the cases when $a = -\infty$ or $b = +\infty$ or both), is an open set. In this section we shall establish a strengthened variant of the converse assertion.

Lemma II.4.1. *If a nonempty set F is closed and bounded above (that is the set of numbers corresponding to the points of the set F is bounded above) it contains the rightmost point.*

A similar assertion applies to a closed set bounded below (which must contain the leftmost point).

Proof. Interpreting the points of F as numbers we put $x_0 = \sup F$. The set F being bounded above, x_0 is a finite number. The definition of supremum readily shows that for any natural number m there is a point $x_m \in F$ such that $x_0 - 1/m < x_m \leq x_0$. Then $x_m \rightarrow x_0$, and, consequently, due to the closedness of F , we have $x_0 \in F$; hence x_0 is the rightmost point of F .

In particular, if a nonempty set F is closed and bounded it contains both the rightmost and the leftmost points. In other words, there is a closed interval $[a, b]$ such that $F \subset [a, b]$ whose end points a and b belong to F . This interval $[a, b]$ is obviously the *smallest* closed interval containing F .

Definition. Let $G \subset R_1$ be an open set. An open interval (α, β) , with $\alpha < \beta$, is called a (connected) *component* of G if $(\alpha, \beta) \subset G$ while the end points α and β do not belong to G . It is allowable here that $\alpha = -\infty$ or $\beta = +\infty$ (or both); for instance, an interval $(\alpha, +\infty)$ is a component of G if it is entirely contained in G and $\alpha \notin G$.

Lemma II.4.2. *Any two components of G have no points in common.*

Proof. Let (α, β) and (γ, δ) be two components of G and let, for instance, $\alpha < \gamma$. We shall show that then $\beta \leq \gamma$,

which will mean that the given intervals do not intersect. Indeed, if we assume that $\beta > \gamma$ we shall have $\alpha < \gamma < \beta$. This will imply that, on the one hand, $\gamma \in (\alpha, \beta) \subset G$ and, on the other hand, γ as an end point of the component (γ, δ) will not belong to G . We thus obtain a contradiction.

Lemma II.4.3. *Every point of an open nonempty set G belongs to one of the components of G .*

Proof. Let $x_0 \in G$. Suppose first that there are points lying on the line to the left and to the right of the point x_0 which do not belong to G and thus belong to its complement $F = \mathbb{R}_1 \setminus G$. Putting

$$F_1 = F \cap [x_0, +\infty]$$

we see that the set F_1 is closed as intersection of two closed sets and is bounded below by x_0 . Consequently, by Lemma II.4.1, F_1 contains the leftmost point β for which we have $\beta > x_0$ (since $x_0 \notin F_1$). Hence, the whole interval $[x_0, \beta) \subset G$. Similarly, there is a point α to the left of x_0 such that $\alpha \notin G$ while $(\alpha, x_0] \subset G$. Taking the union of these intervals $(\alpha, x_0]$ and $[x_0, \beta)$ we obtain an entire interval (α, β) which is a component of G containing the point x_0 .

If there are no points which do not belong to G and lie on one side of x_0 while there are such on the other side we analogously show that x_0 belongs to a component of the type $(\alpha, +\infty)$ or $(-\infty, \beta)$.

Finally, if $G = (-\infty, +\infty)$, i.e. if all the points in the line belong to G , the whole number axis $(-\infty, +\infty)$ is the only component of the set G .

Theorem II.4.1. *Every nonempty open set $G \subset \mathbb{R}_1$ is representable as the union of all its components; the set of the components is at most countable.*

Proof. Let A be the union of all the components of the set G . From Lemma II.4.3 we conclude that $G \subset A$. On the other hand, every component is a subset of G , and therefore $A \subset G$. Consequently, $G = A$.

Let us choose a rational point in each component of G . This sets up a correspondence between the components and the chosen rational numbers. Lemma II.4.2 shows that this correspondence is one-to-one. Since any set of rational numbers is at most countable, so is the collection of the components of G .

Remark. If G is bounded all its connected components are of finite length. If otherwise, there can be (but not necessarily!) one or two infinite intervals among its components.

Theorems II.3.3 and II.4.1 imply that every closed set $F \subset R_1$ is obtained from R_1 by deleting a finite or countable system (which may also be empty if $F = R_1$) of pairwise disjoint open intervals.

In the case when F is bounded and $[a, b]$ is the smallest closed interval containing F the complement $G = R_1 \setminus F$ consists of a system of open intervals contained in (a, b) and of the two semi-infinite intervals $(-\infty, a)$ and $(b, +\infty)$. It follows that in this case F can also be obtained from $[a, b]$ by deleting from the latter an at most countable system of pairwise disjoint open intervals.* Thus, *the complement of a nonempty bounded closed set with respect to the smallest closed interval containing it is an open set.*

§ 5. The Structure of an Open Set in the n -dimensional Space R_n

For open sets in a Euclidean space of dimension $n > 1$ it is impossible to obtain a theorem on the structure of an open set completely analogous to Theorem II.4.1. In this section we shall show that every open nonempty set in R_n can be represented as the union of an at most countable system of open spheres which, generally speaking, can intersect each other; this result is only partially similar to that obtained for open sets in the space R_1 .

Lemma II.5.1. *Let $G \subset R_n$ be a nonempty open set distinct from the whole space R_n and let $x \in G$. Then among all the open spheres with centre at x , contained in G , there exists one of the greatest possible radius.*

Proof. Let E be the set of all numbers r such that $r > 0$ and the open sphere $S(x, r)$ belongs to G . Since x is an interior point of the set G such spheres exist, and hence $E \neq \emptyset$. On the other hand, since $G \neq R_n$, these spheres cannot have an arbitrarily large radius. Consequently, the set E is bounded above. Let us put $r_0 = \sup E$ and show that $S(x, r_0) \subset$

* This system of intervals can also be empty.

$\subset G$. This will of course mean that $S(x, r_0)$ is the sphere whose existence we intend to prove.

Indeed, let $y \in S(x, r_0)$, that is $\rho(y, x) < r_0$. Then there is $r \in E$ such that $\rho(y, x) < r$. Since $r \in E$ we have $S(x, r) \subset G$. But $y \in S(x, r)$ and therefore $y \in G$. Thus, the whole sphere $S(x, r_0)$ is contained in G .

Theorem II.5.1. *Every nonempty open set $G \subset \mathbf{R}_n$ decomposes into a finite or countable system of open spheres (as has been noted, the spheres entering into such a system may intersect).*

Proof. If $G = \mathbf{R}_n$ we simply put $G = \bigcup_{m=1}^{\infty} S(O, m)$ where O is the origin.* Therefore we shall suppose that $G \neq \mathbf{R}_n$.

Let A be the set of all rational points contained in G . As will be seen, A is nonempty, and therefore it is clear that it is countable; consequently the points of A can be numbered $x_1, x_2, \dots, x_m, \dots$. By the lemma, for each $x_m \in A$ there exists the open sphere (denote it by S_m) of maximal radius with centre at x contained in G . Let us prove that $G = \bigcup_{m=1}^{\infty} S_m$ (in particular, this will show that A is nonempty).

Obviously we should only check that $G \subset \bigcup_{m=1}^{\infty} S_m$. Let $y \in G$. Then $S(y, \varepsilon) \subset G$ for some $\varepsilon > 0$. Let us choose

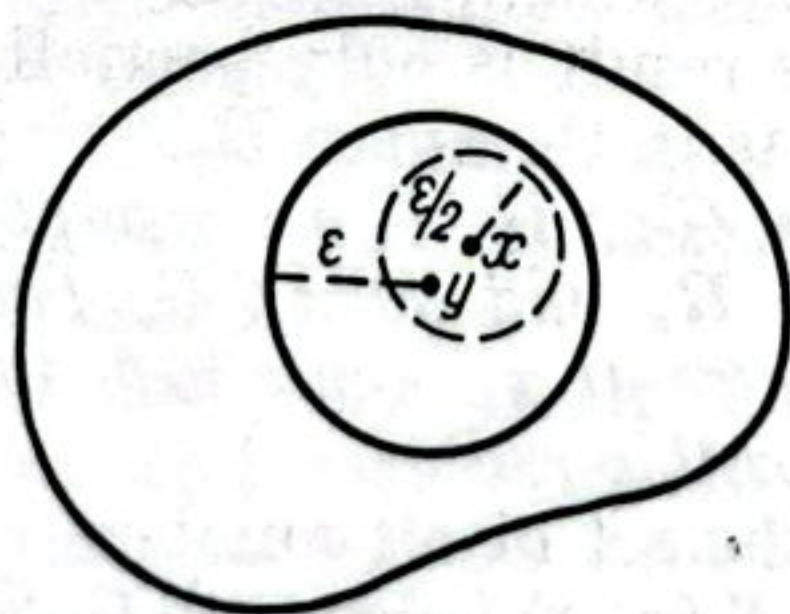


Fig. 4

* If the space $\mathbf{R}_n$ itself were thought of as a sphere of an infinite radius (with centre at any point) the case $G = \mathbf{R}_n$ would be trivial since G would itself be a sphere. But in our course, when speaking of spheres, we always mean those of finite radius.

a rational point $x \in R_n$ such that $\rho(x, y) < \varepsilon/2$. Then, moreover, we have $x \in S(y, \varepsilon) \subset G$, and, consequently, x is a point belonging to the set A . Let $x = x_k$. We have $S(x, \varepsilon/2) \subset G$. Indeed, if $z \in S(x, \varepsilon/2)$ then $\rho(z, y) \leq \rho(z, x) + \rho(x, y) < \varepsilon/2 + \varepsilon/2 = \varepsilon$, i.e. $z \in S(y, \varepsilon) \subset G$ (Fig. 4). The inclusion relation $S(x_k, \varepsilon/2) \subset G$ shows that the sphere S_k must be of radius not smaller than $\varepsilon/2$, and therefore $y \in S_k$. Thus, by the arbitrariness of the point y chosen from the set G , we see that $G \subset \bigcup_{m=1}^{\infty} S_m$.*

Another representation of open sets in the n -dimensional space R_n will be studied in Chapter V.

§ 6. Finite Covering Theorem

The theorem proved in this section plays an important role in many aspects of the theory of functions.

Definition. Let $A \subset R_n$. A system $\mathcal{G}$ of sets $G_\alpha \subset R_n$ is called a *covering* of the set A if any point $x \in A$ belongs to at least one of the sets $G_\alpha \in \mathcal{G}$. If, in addition, all the sets G_α are open the covering $\mathcal{G}$ itself is called *open*.

Theorem II.6.1. (Borel-Lebesgue**). *Every open covering $\mathcal{G}$ of a bounded closed set $F \subset R_n$ contains a finite subcovering of F , that is a finite number of sets $G_\alpha \in \mathcal{G}$ which also form a covering of the set F .****

Proof. To simplify the notation we shall consider the case $n = 2$ (the general case of an arbitrary n is treated quite similarly).

Suppose that it is impossible to choose a finite subsystem of $\mathcal{G}$ which covers the set F . Since F is bounded it is contain-

* In the course of the proof we have shown that G is representable as a countable union of open spheres. But it is clear that there can also be the case when a given open set G is a union of only a finite number p of open spheres.

** F.É.É. Borel (1871-1956) and H. Lebesgue (1875-1941), prominent French mathematicians, contributed many fundamental results to the theory of functions of a real variable.

*** This theorem describes the so-called *Heine-Borel property* connected with the general definition of the notion of *compactness* playing an extremely important role in functional analysis (e.g. see K. Yosida, *Functional Analysis*, Springer, 1966).—Tr.

ed within a closed rectangle $\Delta = [a, b; c, d]$.^{*} Let us divide this rectangle into four equal parts by straight lines parallel to the coordinate axes.^{**} Then among the closed rectangles thus obtained there is at least one which contains a part of the set F for which it is impossible to choose from $\mathcal{G}$ a finite subcovering.^{***} Let this rectangle be $\Delta_1 = [a_1, b_1; c_1, d_1]$. Then

$$b_1 - a_1 = \frac{b - a}{2}, \quad d_1 - c_1 = \frac{d - c}{2}$$

The rectangle Δ_1 is again divided into four equal parts from which we choose one (denote it Δ_2) containing a part of the set F which cannot be covered by a finite subsystem of $\mathcal{G}$. Proceeding in this way indefinitely we arrive at a sequence of rectangles $\Delta_m = [a_m, b_m; c_m, d_m]$ where

$$b_m - a_m = \frac{b - a}{2^m}, \quad d_m - c_m = \frac{d - c}{2^m}$$

such that $\Delta_{m+1} \subset \Delta_m$ for all m and each Δ_m containing a part of the set F for which there exists no finite covering consisting of sets belonging to $\mathcal{G}$. In particular, it follows that each Δ_m contains an infinite number of points belonging to F . By the well-known *nested interval theorem*, the intervals $[a_m, b_m]$ contract to a common point $\xi_1^{(0)}$ and the intervals $[c_m, d_m]$ to a common point $\xi_2^{(0)}$. Then the point $x_0 = (\xi_1^{(0)}, \xi_2^{(0)})$ belonging to R_2 is contained in Δ_m for all m .

Let us show that x_0 is a limit point of the set F . Let S be an ε -neighbourhood of the point x_0 where $\varepsilon > 0$ is taken arbitrarily. Then there is m such that

$$b_m - a_m < \frac{\varepsilon}{\sqrt{2}} \quad \text{and} \quad d_m - c_m < \frac{\varepsilon}{\sqrt{2}}$$

^{*} In this way we symbolically denote the rectangle specified by the inequalities $a \leq \xi_1 \leq b$ and $c \leq \xi_2 \leq d$.

^{**} In the case $n = 3$ we deal with a rectangular parallelepiped Δ which should be divided, by means of planes, into 8 equal parts. In the general case of the space R_n the corresponding figure is divided into $2^{(n)}$ parts.

^{***} If there are several rectangles possessing this property we arbitrarily choose one of them. It should be noted that if for each of the four parts of the set F there existed a finite subsystem of $\mathcal{G}$ covering F there would exist a finite subcovering for the whole set F as well.

The diagonal of the rectangle Δ_m is therefore less than ε and hence $\Delta_m \subset S$.^{*} Consequently, together with Δ_m , the neighbourhood S contains an infinitude of points of F , whence, by the arbitrariness of ε , it readily follows that x_0 is a limit point of F .

Since F is closed, $x_0 \in F$. Hence, $x_0 \in G_\alpha$ for some α ($G_\alpha \in \mathfrak{G}$). Therefore there exists $\varepsilon > 0$ such that $S(x_0, \varepsilon) \subset$

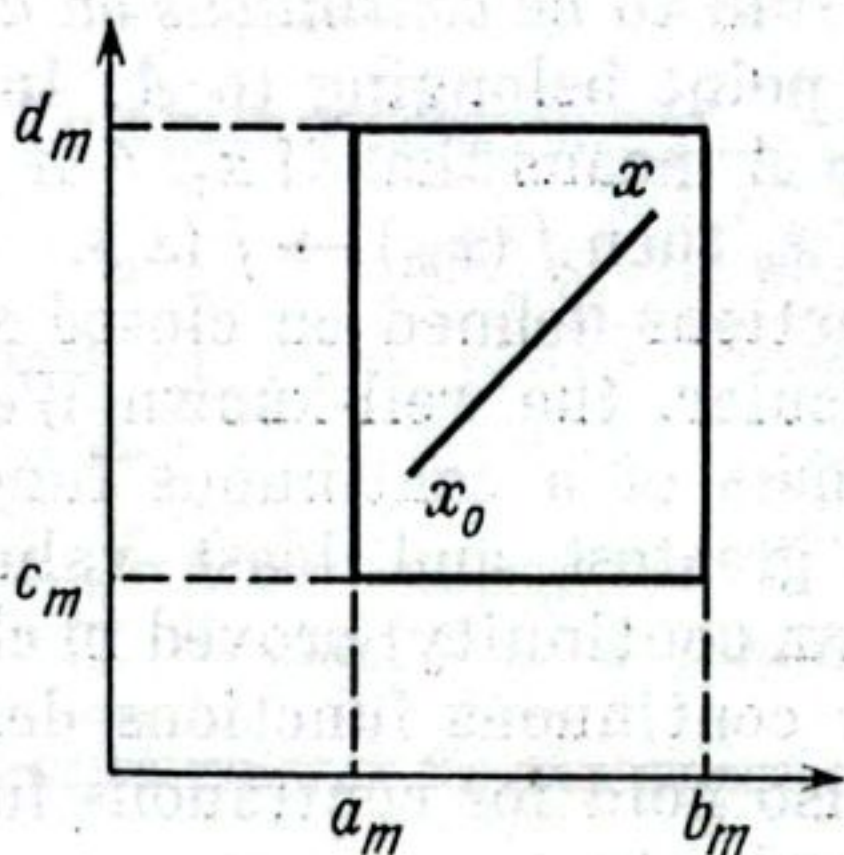


Fig. 5

$\subset G_\alpha$. If now we choose m so that $\Delta_m \subset S(x_0, \varepsilon)$ the whole rectangle Δ_m , together with the part of the set F it contains, will be covered by a system consisting of one of the sets G_α . This contradicts the construction of the rectangles Δ_m and thus proves the theorem.

§ 7. Continuous Functions

Every function f of n real variables $\xi_1, \xi_2, \dots, \xi_n$ can be interpreted as a function $f(x)$ of the point $x = (\xi_1, \xi_2, \dots, \xi_n)$ belonging to the space R_n . Let us consider a real function f defined on an arbitrary set $A \subset R_n$. Let $x_0 \in A$. As usual, the function f is said to be *continuous at the point* x_0 if $f(x_m) \rightarrow f(x_0)$ for any sequence of points $x_m \in A$ convergent to x_0 .

^{*} In this proof it is important that the distance from x_0 to any point $x \in \Delta_m$ is less than ε (Fig. 5). In the n -dimensional case the argument is similar but $\varepsilon/\sqrt{2}$ is replaced by $\varepsilon/\sqrt{n}$.

In this definition we do not exclude the case when x_0 is an isolated point of the set A . But, if x_0 is an isolated point of A and $\{x_m\}$ is an arbitrary sequence such that $x_m \rightarrow x_0$ ($x_m \in A$) then, beginning with some value of m , we must have $x_m = x_0$, and the relation $f(x_m) \rightarrow f(x_0)$ is fulfilled automatically. Thus, *every function is continuous at any isolated point of the set on which it is defined.*

A function f is said to be *continuous on a set A* if it is continuous at every point belonging to A . In other words, the continuity of f on A means that if $x_m \in A$ ($m = 1, 2, \dots$), $x_0 \in A$ and $x_m \rightarrow x_0$ then $f(x_m) \rightarrow f(x_0)$.

Continuous functions defined on closed sets are of special interest. In particular, the well-known *Weierstrass' theorems* (on the boundedness of a continuous function and on the existence of its greatest and least values) and *Cantor's theorem* (on uniform continuity) proved in classical mathematical analysis for continuous functions defined on bounded closed domains also hold for continuous functions on arbitrary bounded closed sets.

In what follows we shall frequently use the symbol $A[C]$, where A is a given set and by C is meant a certain property or condition; this symbol will be understood as denoting the subset of A at whose all points the given condition or property holds. For instance, if f is a function defined on A and a is a real number the symbol $A[f(x) \geq a]$ denotes the set of all $x \in A$ for which $f(x) \geq a$.

Theorem II.7.1. *For a function f defined on a closed set F to be continuous on that set it is necessary and sufficient that the sets $F[f(x) \geq a]$ and $F[f(x) \leq a]$ be closed for any real a .*

Proof. (a) *Necessity.* Let a be chosen and $E = F[f(x) \geq a]$. Taking an arbitrary sequence of points $x_m \in E$ having a limit ($x_m \rightarrow x_0$) we conclude that, since F is closed, the point x_0 must belong to F and that, since f is continuous, $f(x_m) \rightarrow f(x_0)$. But $f(x_m) \geq a$ for all m , and we also have $f(x_0) \geq a$, that is $x_0 \in E$. Thus, E is closed. The closedness of the set $F[f(x) \leq a]$ is shown analogously.

(b) *Sufficiency.* Let $x_m \in F$ ($m = 1, 2, \dots$) and $x_m \rightarrow x_0$ (then also $x_0 \in F$). Choosing an arbitrary $\varepsilon > 0$ we put

$$E_1 = F[f(x) \geq f(x_0) + \varepsilon], \quad E_2 = F[f(x) \leq f(x_0) - \varepsilon]$$

By the hypothesis, both sets are closed, and therefore their union $E = E_1 \cup E_2$ is also closed. On the other hand, $x_0 \notin \bar{E}$, and hence x_0 cannot be a limit point of the set E either. According to Theorem II.2.1, this means that there is a neighbourhood $S(x_0, \delta)$ containing no points of the set E . But we have $x_m \in S(x_0, \delta)$ for $m \geq M$, and thus $x_m \in E$ for $m \geq M$ whence

$$f(x_0) - \varepsilon < f(x_m) < f(x_0) + \varepsilon$$

By the arbitrariness of ε , it follows that $f(x_m) \rightarrow f(x_0)$. The continuity of f has thus been established.

Corollary. *If F_i ($i = 1, 2, \dots, k$) are closed pairwise disjoint sets, $F = \bigcup_{i=1}^k F_i$ and f is a function defined on F such that it is constant on each F_i , the function f is continuous on F .*

Indeed, by Theorem II.3.1, the set F is closed. The function f being constant on every F_i , each of the sets $F[f(x) \geq a]$ and $F[f(x) \leq a]$ is closed for any a since it is either a union of some F_i 's or an empty set.

Remark. If the function f is defined throughout $\mathbf{R}_n$ the above continuity criterion can also be stated as follows: *for f to be continuous throughout $\mathbf{R}_n$ it is necessary and sufficient that the sets $\mathbf{R}_n[f(x) < a]$ and $\mathbf{R}_n[f(x) > a]$ be open for any real a .* Indeed, these sets are, respectively, the complements of the sets $\mathbf{R}_n[f(x) \geq a]$ and $\mathbf{R}_n[f(x) \leq a]$ with respect to the whole space $\mathbf{R}_n$, and therefore the condition that the former sets are open is equivalent to the condition that the latter are closed. But it should be noted that for a function defined on an arbitrary closed set F only the first (basic) statement of Theorem II.7.1 holds.

§ 8. Borel Sets

In this section we shall consider a certain class of subsets of the space $\mathbf{R}_n$. Using the notions introduced in Sec. I.7 we state the following

Definition. A set $A \subset \mathbf{R}_n$ is called a *Borel set* if it belongs

to the σ -algebra generated by the system of all closed subsets of R_n .*

The collection of all Borel sets in R_n will be denoted by $\mathfrak{B}$. According to the definition, every closed set F belongs then to $\mathfrak{B}$; also every open set $G \in \mathfrak{B}$ because the open sets are the complements of the closed sets. But the class $\mathfrak{B}$ also includes many other types of sets of a more complicated structure which are neither open nor closed. For instance, every set representable as a countable union of closed sets (such a set is spoken of as an F_σ -set) is a Borel set. In particular, any countable set is representable as the union of single-element sets which are closed and is therefore a Borel F_σ -set. Similarly, every set representable as the intersection of a countable number of open sets is also a Borel set (spoken of as a G_δ -set). But the totality of all F_σ - and G_δ -sets does not exhaust the class $\mathfrak{B}$ of all Borel sets either.

It is obvious that the σ -algebra generated by the family of all open sets in R_n coincides with $\mathfrak{B}$. We shall also prove a stronger assertion.

Theorem II.8.1. *The σ -algebra generated by the system of all open spheres $S \subset R_n$ coincides with $\mathfrak{B}$.*

Proof. Let us denote temporarily the σ -algebra generated by the collection of all open spheres by $\mathfrak{B}_1$. From Theorem II.5.1 it follows that every open set in R_n belongs to $\mathfrak{B}_1$ and therefore every closed set in R_n is also contained in $\mathfrak{B}_1$. But $\mathfrak{B}$ is the minimal σ -algebra containing all the closed sets; consequently, $\mathfrak{B} \subset \mathfrak{B}_1$. On the other hand, all the open spheres S belong to $\mathfrak{B}$ and therefore $\mathfrak{B}_1 \subset \mathfrak{B}$. Thus $\mathfrak{B}_1 = \mathfrak{B}$, which is what we set out to prove.

* Since the whole space R_n is closed the class of all Borel sets can be defined as the σ -ring generated by the family of all closed sets.

Metric Spaces

§ 1. The Definition of a Metric Space

In the previous chapter we considered a simple generalization of the concept of the three-dimensional (geometric) space studied in elementary mathematics: we introduced the notion of an n -dimensional Euclidean space. In modern mathematics the concept of space undergoes further generalization: by a space is meant an abstract set of arbitrary objects (e.g. collections of numbers, functions and families of functions, some geometrical objects, etc.) for which some relationships are introduced analogous to spatial relationships between geometrical objects in the ordinary three-dimensional space. In particular, in this section we shall deal with arbitrary sets consisting of arbitrary elements on which only one requirement is imposed that there should be defined the notion of the distance (between any two elements) satisfying some conditions (axioms)*. This leads to the concept of a metric space whose strict definition is given below.

Definition. By a *metric space* we mean every set E in which to any two elements $x, y \in E$ a real nonnegative number $\rho(x, y)$ is assigned so that the function $\rho(x, y)$ thus defined on the totality of all pairs of elements x, y (called a *distance* or *distance function* or *metric*) satisfies the following requirements (*axioms of metric* or *axioms of metric space*):

- (i) $\rho(x, y) = 0$ if and only if $x = y$;
- (ii) $\rho(x, y) = \rho(y, x)$ (*symmetry*);
- (iii) $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$ for any three elements x, y , and z (*triangle inequality*).

* In our course metric spaces will be the most general type of spaces but modern mathematics deals with still more general spaces, e.g. topological spaces.

The elements of a metric space are more often called its *points*.

In Sec. II.1 we saw that the distance between the points in the n -dimensional Euclidean space R_n possesses all these properties. Hence, R_n is a special case of a metric space. In the general definition of a metric space given above all these properties are introduced as axioms and are not deduced from other facts related to the concrete structure of the space in question.

Let us consider some examples of metric spaces.

1. Let $\mathcal{I}$ be the set consisting of all infinite number sequences $x = (\xi_1, \xi_2, \dots, \xi_i, \dots)$ satisfying the condition

$$\sum_{i=1}^{\infty} |\xi_i| < +\infty$$

Here we suppose that all ξ_i 's are real numbers*. Taking two arbitrary elements $x = \{\xi_i\}$ and $y = \{\eta_i\}$ belonging to $\mathcal{I}$ we put

$$\rho(x, y) = \sum_{i=1}^{\infty} |\xi_i - \eta_i|$$

The fulfilment of the first two axioms of metric space is quite obvious for the function $\rho(x, y)$ thus introduced. The triangle inequality is also easily established, for, given arbitrary $x = \{\xi_i\}$, $y = \{\eta_i\}$ and $z = \{\zeta_i\}$, we have

$$\begin{aligned} \rho(x, y) = \sum_{i=1}^{\infty} |\xi_i - \eta_i| &\leq \sum_{i=1}^{\infty} |\xi_i - \zeta_i| + \\ &+ \sum_{i=1}^{\infty} |\zeta_i - \eta_i| = \rho(x, z) + \rho(z, y) \end{aligned}$$

Thus, $\mathcal{I}$ is a metric space.

* In this example as well as in the following ones we must not necessarily limit ourselves to sequences of real numbers; we can similarly consider analogous complex spaces of sequences consisting of complex numbers.

2. Let us denote by l^2 the set of all sequences $x = \{\xi_i\}$ of real numbers for which $\sum_{i=1}^{\infty} \xi_i^2 < +\infty$; putting

$$\rho(x, y) = \sqrt{\sum_{i=1}^{\infty} (\xi_i - \eta_i)^2}$$

we obtain the metric space l^2 .

To prove that l^2 is a metric space we must first of all show that $\rho(x, y)$ is a function assuming only finite values, that is we have to prove the convergence of the series entering into the right-hand side for any x and y belonging to l^2 . To this end we shall begin with generalizing Cauchy's inequality (4) in Chapter II established for finite sequences to the case of infinite sequences of numbers a_i, b_i ($i = 1, 2, \dots$). Indeed, taking an arbitrary natural number n we can write down inequality (4) (Chapter II) for this n and then pass to the limit for $n \rightarrow \infty$, which results in the inequality

$$\sqrt{\sum_{i=1}^{\infty} (a_i + b_i)^2} \leq \sqrt{\sum_{i=1}^{\infty} a_i^2} + \sqrt{\sum_{i=1}^{\infty} b_i^2} \quad (1)$$

We shall call (1) *Cauchy's inequality for infinite number sequences*. From inequalities (3) established in Chapter II we similarly derive the other Cauchy inequality for infinite sequences:

$$\left(\sum_{i=1}^{\infty} a_i b_i\right)^2 \leq \left(\sum_{i=1}^{\infty} a_i^2\right) \left(\sum_{i=1}^{\infty} b_i^2\right) \quad (2)$$

In particular, from inequality (1) it follows that if $x = \{\xi_i\} \in l^2$ and $y = \{\eta_i\} \in l^2$ the sequence $\{\xi_i - \eta_i\}$ also belongs to l^2 , that is $\rho(x, y) < +\infty$.

Now the verification of the axioms of metric space for l^2 is carried out in the complete analogy with Section II.1 where this was done for R_n .

The metric space l^2 is sometimes referred to as an *infinite-dimensional Euclidean space*.

3. Let m be the space consisting of all bounded infinite number sequences with distance function defined for $x = \{\xi_i\}$ and $y = \{\eta_i\}$ by the formula

$$\rho(x, y) = \sup_i |\xi_i - \eta_i|$$

Let the reader show that m is in fact a metric space.

By analogy with R_n , we shall call the numbers ξ_i , in the sequence $x = \{\xi_i\}$ serving as points of metric spaces of number sequences, the *coordinates* of the point x .

4. Let C be the set of all real continuous functions $x(t)$ defined on an interval $a \leq t \leq b$, the distance $\rho(x, y)$ between any two elements $x, y \in C$ being defined by the formula

$$\rho(x, y) = \max |x(t) - y(t)|$$

We see that the distance in C is the maximal deviation of one function from another.

It is clear that $\rho(x, y) = 0$ if and only if $x(t) \equiv y(t)$; the axiom of symmetry is also obvious. The triangle inequality for any three functions x, y and z belonging to C is checked as follows:

$$\begin{aligned} |x(t) - y(t)| &\leq |x(t) - z(t)| + |z(t) - y(t)| \leq \\ &\leq \max |x(t) - z(t)| + \max |z(t) - y(t)| = \\ &= \rho(x, z) + \rho(z, y) \end{aligned}$$

Since this is true for any $t \in [a, b]$ the same inequality holds, in particular, for $\rho(x, y) = \max |x(t) - y(t)|$.

Thus, C is a metric space.

The distance function can be introduced similarly in the set of all continuous functions defined on an arbitrary bounded closed set in the space R_n , which turns this set of functions into a metric space.

5. Let us take the same set of functions as in the foregoing example but define the distance with the aid of another formula:

$$\rho(x, y) = \int_a^b |x(t) - y(t)| dt$$

It can readily be checked that all the axioms of metric space hold in this case. Hence, we deal with a metric space consisting of the same elements as C but having a different metric. Since the definition of a metric space includes both the indication of the set of elements and the indication of the distance the latter metric space (we shall denote it C_L)

is different from the space C although it consists of the same elements.

If we are given a set the operation of introducing (defining) a distance function for its elements satisfying all the axioms of metric is spoken of as *metrization*. We can thus say that in Examples 4 and 5 we have metrized the set of continuous functions in two different ways.

It should be noted that every set can be metrized trivially if we put, for instance,

$$\rho(x, y) = \begin{cases} 1 & \text{for } x \neq y \\ 0 & \text{for } x = y \end{cases}$$

The axioms of metric space obviously hold for such a distance function which is of course of no practical interest.

The concept of a metric space was introduced at the beginning of the 20th century by M. R. Fréchet, a French mathematician.

§ 2. Convergence in a Metric Space

The distance function in an arbitrary metric space E makes it possible to introduce the notion of the limit by analogy with construction used in the previous chapter in the case of the space R_n . Namely, we shall write $x_n \rightarrow x$ if $\rho(x_n, x) \rightarrow 0$. This type of convergence of a sequence of points will be spoken of as *convergence with respect to the distance* or *with respect to the metric of the given space E* (or *convergence in the metric space E*).

Let us prove that a sequence of points of a metric space can only possess a unique limit (provided it exists). Indeed, let $x_n \rightarrow x$ and $x_n \rightarrow y$. By the triangle inequality, $\rho(x, y) \leq \rho(x, x_n) + \rho(x_n, y)$. The right-hand side of this inequality tends to zero as $n \rightarrow \infty$ while the left-hand side is nonnegative, consequently $\rho(x, y) = 0$, that is $x = y$.

Let us prove one more inequality which holds for any four points x, y, z and u belonging to E :

$$|\rho(x, y) - \rho(z, u)| \leq \rho(x, z) + \rho(y, u) \quad (3)$$

Indeed, the triangle inequality implies

$$\rho(x, y) \leq \rho(x, z) + \rho(z, y) \leq \rho(x, z) + \rho(z, u) + \rho(y, u)$$

whence

$$\rho(x, y) - \rho(z, u) \leq \rho(x, z) + \rho(y, u) \quad (4)$$

In the last inequality x and y can be, respectively, interchanged with z and u , which results in

$$\rho(z, u) - \rho(x, y) \leq \rho(x, z) + \rho(y, u) \quad (5)$$

Now, inequalities (4) and (5) imply (3).

As a consequence of inequality (3) we can state that *the distance $\rho(x, y)$ is a continuous function of x and y in the sense that if $x_n \rightarrow x$ and $y_n \rightarrow y$ then $\rho(x_n, y_n) \rightarrow \rho(x, y)$.*

Indeed, by (3) we can write

$$|\rho(x_n, y_n) - \rho(x, y)| \leq \rho(x_n, x) + \rho(y_n, y) \rightarrow 0$$

It should also be noted that if $x_n \rightarrow x$ any subsequence of points x_{n_k} ($n_1 < \dots < n_k < \dots$) of the sequence $\{x_n\}$ is also convergent to x . If, beginning with some n_0 , $x_n = x$ for all $n \geq n_0$ then $\lim x_n = x$.

Let us elucidate the concrete meaning of the convergence for the metric spaces in the above example.

From the definitions of the distance functions in l and l^2 it follows that in these spaces *the convergence with respect to distance implies the coordinate-wise convergence*: if $x_n = \{\xi_i^{(n)}\} \rightarrow x = \{\xi_i\}$ then $\xi_i^{(n)} \rightarrow \xi_i$ for each $i = 1, 2, \dots$. The converse conclusion is not true, that is, in contrast to $\mathbf{R}_n$, the convergence with respect to distance in l and in l^2 is a stronger condition than the simple coordinate-wise convergence. Indeed, let x_n be a point whose n th coordinate is equal to 1 while all the other coordinates are equal to 0 and let x be "the zero point" whose all coordinates are equal to zero. Then the sequence $\{x_n\}$ is obviously convergent to x with respect to coordinates but at the same time $\rho(x_n, x) = 1$ for all n .

In the space m the convergence $x_n \rightarrow x$ is nothing but the *uniform coordinate-wise convergence*, that is $x_n \rightarrow x$ implies that $\xi_i^{(n)} \rightarrow \xi_i$ ($i = 1, 2, \dots$) uniformly with respect to i . *The convergence in the space C (i.e. with respect to its metric) is equivalent to the uniform convergence of the functional sequence*, that is $x_n \rightarrow x$ ($x_n, x \in C$) means that

$$\max |x_n(t) - x(t)| \rightarrow 0$$

The convergence in C_L imposes essentially weaker requirements on the functional sequence: a sequence of elements belonging to C_L convergent with respect to the metric of C_L may even be divergent, as functional sequence, at each point. Here we shall confine ourselves to a simpler example of a sequence of continuous functions defined on the interval $[0, 1]$ convergent with respect to the distance in C_L to the

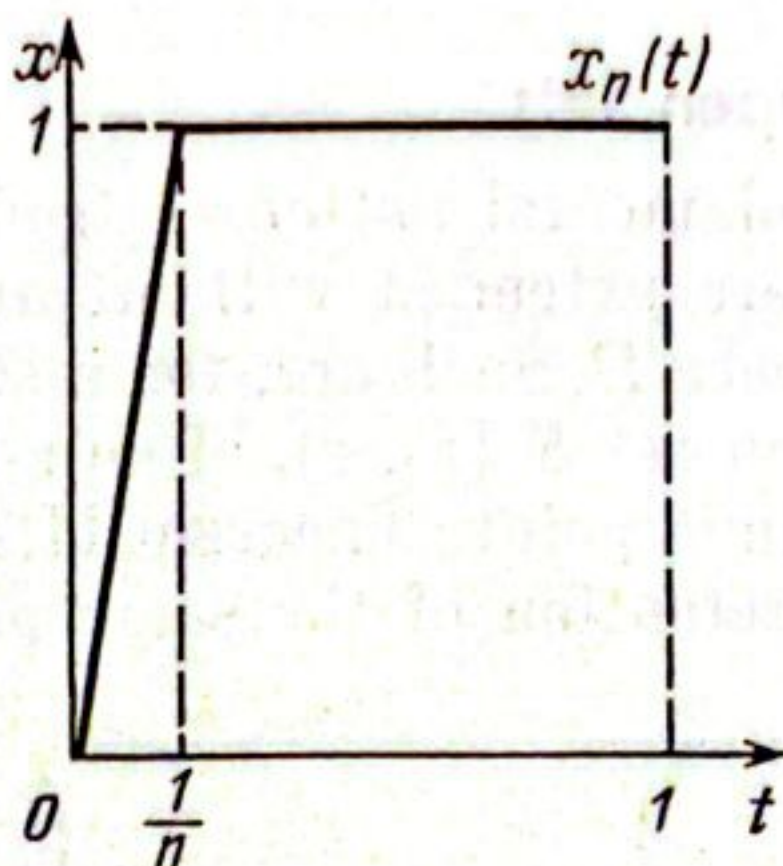


Fig. 6

function $x(t) \equiv 1$ and not convergent to unity at the point $t = 0$. Namely, we put

$$x_n(t) = \begin{cases} nt & \text{for } 0 \leq t \leq \frac{1}{n} \\ 1 & \text{for } \frac{1}{n} \leq t \leq 1 \end{cases}$$

(see Fig. 6). It can readily be shown by means of direct computations that for $x(t) \equiv 1$ we have

$$\rho(x_n, x) = \frac{1}{2n} \rightarrow 0$$

with respect to the distance in the space C_L while $\lim_{n \rightarrow \infty} x_n(0) = 0 \neq 1$.

For any two functions f and g defined on an interval $[a, b]$ the number

$$\frac{1}{b-a} \int_a^b |f(t) - g(t)| dt$$

is called the *mean deviation of order one* of one of these functions from another. It is evident that the convergence in C_L is equivalent to the condition that the corresponding sequence of mean deviations tends to zero.

It is readily seen that if a sequence of continuous functions converges to a limit in C it also converges to the same limit in C_L . As was shown above, the converse is not true.

§ 3. Closed and Open Sets

A number of fundamental notions introduced in Sec. II.2 for the space R_n are extended without any changes to an arbitrary metric space E . Such are, for instance, the notions of an open ball (sphere) $S(x_0, \varepsilon)$, of an ε -neighbourhood of a point and of a limit point. Theorem II.2.1 also continues to hold (with the retention of the same proof). Besides the

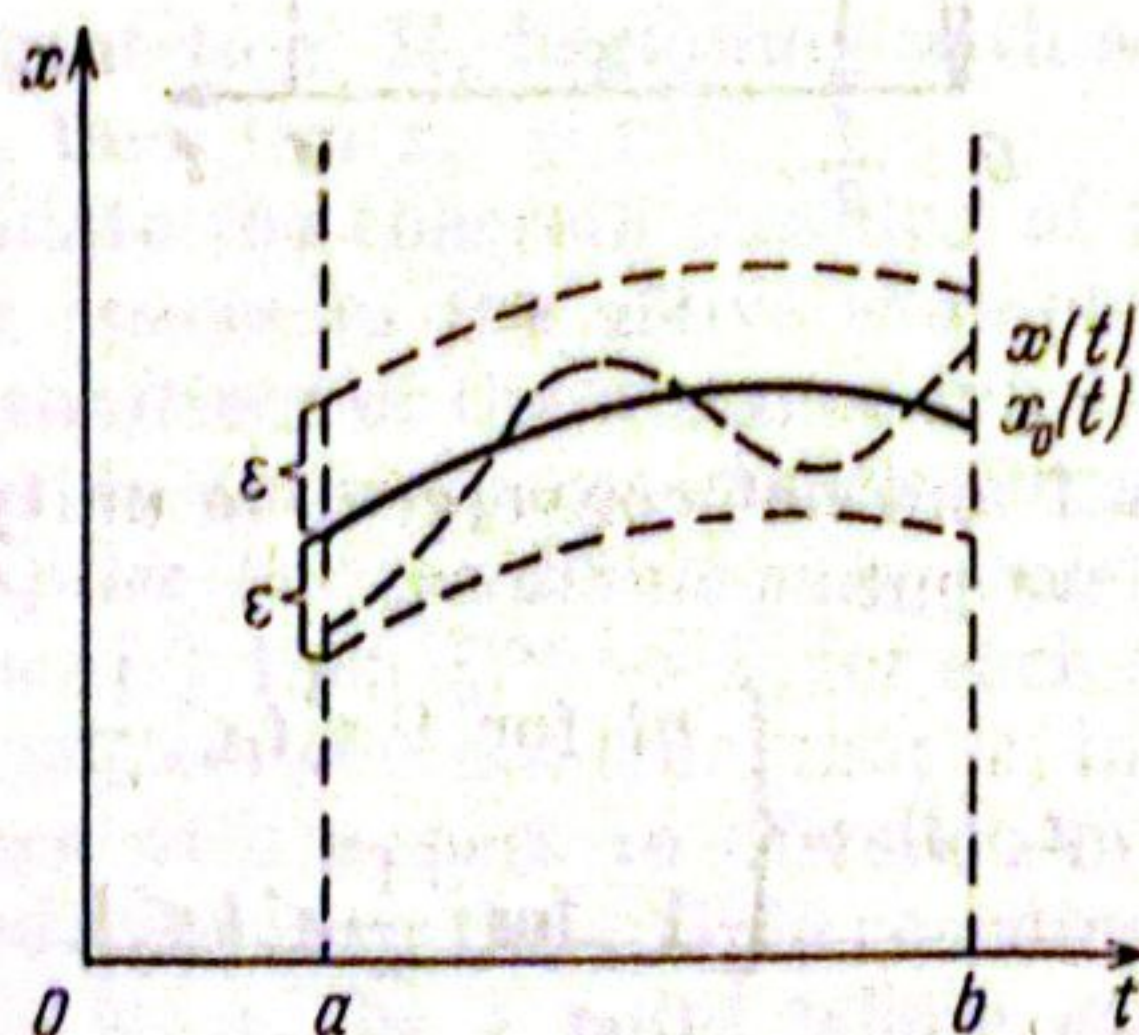


Fig. 7

notion of an open sphere it is advisable to introduce that of a *closed sphere (ball)* $S^*(x_0, \varepsilon)$ (see the footnote on p. 43). It is defined as the set of all points $x \in E$ for which $\rho(x, x_0) \leq \varepsilon$.

The closed sphere $S^*(x_0, \varepsilon)$ in R_1 is nothing but the closed interval $[x_0 - \varepsilon, x_0 + \varepsilon]$. In the space C the ball $S^*(x_0, \varepsilon)$ consists of all functions x satisfying the condition $|x(t) - x_0(t)| \leq \varepsilon$ throughout the interval $[a, b]$ on which the functions belonging to C are defined, that is the functions whose graphs lie within the strip shown in Fig. 7.

The definition of a *closed set* is stated for an arbitrary metric space precisely by analogy with the case of the space R_n . In particular, the whole space E , the empty set $\emptyset$ and any finite set (consisting of a finite number of points belonging to E) are closed.

Let us show that the closed sphere $S^*(x_0, r)$ is a closed set.

To this end, suppose that $x_n \in S^*(x_0, r)$; then $\rho(x_n, x_0) \leq r$. If $x_n \rightarrow x$ the continuity of the distance function shows that

$$\rho(x, x_0) = \lim \rho(x_n, x_0) \leq r$$

and, consequently, x also belongs to $S^*(x_0, r)$.

The basic properties of closed sets established for the space R_n in Theorems II.3.1-2 turn out to hold for an arbitrary metric space as well.

The notion of a closed set is related to the operation of adding to an arbitrary set $A \subset E$ all the limits of convergent sequences of points belonging to A . The resultant set, denoted $\bar{A}$, is called the *closure* of the set A .

Thus, we always have $A \subset \bar{A}$. A set A is closed if and only if $A = \bar{A}$.

Let us prove that in the space R_n the equality

$$\overline{S(x_0, r)} = S^*(x_0, r) \quad (6)$$

holds for any open sphere $S(x_0, r)$.

From the continuity of the distance it readily follows that no point lying outside $S^*(x_0, r)$ can be a limit point of the ball $S(x_0, r)$. On the other hand, let $x = (\xi_1, \xi_2, \dots, \xi_n)$ be a point such that $\rho(x, x_0) = r$. Let us put

$$x_m = (\xi_1^{(0)} + t_m(\xi_1 - \xi_1^{(0)}), \dots, \xi_n^{(0)} + t_m(\xi_n - \xi_n^{(0)})) \\ (m = 1, 2, \dots)$$

where $\xi_i^{(0)}$ ($i = 1, 2, \dots$) are the coordinates of the point x_0 and $t_m = 1 - \frac{1}{m}$. Then $x_m \in S(x_0, r)$ and $x_m \rightarrow x$; consequently $x \in \overline{S(x_0, r)}$.

These considerations immediately imply equality (6)*.

* It should be stressed that for an arbitrary metric space we can only guarantee the inclusion relation $\overline{S(x_0, r)} \subset S^*(x_0, r)$; in the general case these sets do not necessarily coincide.

Let $A \subset E$ where E is again an arbitrary metric space. If $x \in \bar{A}$ then, given an arbitrary $\varepsilon > 0$, there is a point $x' \in A$ such that $\rho(x', x) < \varepsilon$. Indeed, if $x \in A$ we can simply put $x' = x$ and if $x \notin A$ the inclusion relation $x \in \bar{A}$ means that $x = \lim x_n$ where $x_n \in A$. In this case we can choose as x' a point x_n with a sufficiently large number n .

Theorem III.3.1. *The closure $\bar{A}$ of A is the smallest closed set containing A .*

Proof. Let us first show that $\bar{A}$ is always a closed set. Let $x_n \in \bar{A}$ and $x_n \rightarrow x$. According to the property discussed above, for each n there exists $y_n \in A$ such that $\rho(y_n, x_n) < \frac{1}{n}$. Then

$$\rho(y_n, x) \leq \rho(y_n, x_n) + \rho(x_n, x) \rightarrow 0$$

and consequently $y_n \rightarrow x$. Thus, $x \in \bar{A}$, which means that the set $\bar{A}$ is closed.

Now note that every closed set F containing A also contains all the limits of the convergent sequences whose members belong to A . Hence, $\bar{A} \subset F$. Thus, since the set $\bar{A}$ is itself closed it is the smallest of the closed sets $F \supset A$.

The notion of an interior point and the definition of an open set stated for the space $\mathbf{R}_n$ are readily extended without changes to an arbitrary metric space E . Every open sphere in E is an open set. Theorems II.3.3-5 establishing the relationship between closed and open sets in $\mathbf{R}_n$ and the basic properties of open sets continue to hold in the general case as well.

§ 4. Complete Metric Spaces

The fact that for a sequence of real numbers to be convergent it is not only necessary but also sufficient that the sequence satisfy *Cauchy's condition*, i.e. be *fundamental*, plays an important role in classical mathematical analysis. This property extends completely to the Euclidean space $\mathbf{R}_n$: a sequence of points x_m belonging to $\mathbf{R}_n$ converges to a limit if and only if $\rho(x_m, x_p) \rightarrow 0$ for $m, p \rightarrow \infty$.

Indeed, let $x_m = (\xi_1^{(m)}, \xi_2^{(m)}, \dots, \xi_n^{(m)})$ and let $\rho(x_m, x_p) \rightarrow 0$ for $m, p \rightarrow \infty$. This implies

$$\xi_i^{(m)} - \xi_i^{(p)} \rightarrow 0 \text{ for each } i$$

and therefore, by Cauchy's convergence criterion, the number sequences $\lim \xi_i^{(m)} = \xi_i$ ($i = 1, 2, \dots, n$) possess finite limits. Therefore $x_m \rightarrow x = (\xi_1, \xi_2, \dots, \xi_n)$.

Let the reader prove the "if" part of this proposition. For an arbitrary metric space the necessity of Cauchy's criterion will be proved in Theorem III.4.1 below.

It turns out that for an arbitrary metric space Cauchy's criterion is not sufficient, that is a Cauchy sequence (fundamental sequence) may be divergent. For instance, considering the set of all rational numbers as a metric space with distance defined by the formula used in the case of the space $\mathbf{R}_1$, i.e. $\rho(x, y) = |x - y|$, we readily see that a Cauchy sequence composed of rational numbers may not be convergent in this space if its limit is an irrational number. In this section we consider a class of metric spaces for which the Cauchy convergence criterion continues to be necessary and sufficient.

Definition. A sequence of points x_n belonging to a metric space E is said to be *fundamental* (or to be a *Cauchy sequence*) if $\rho(x_n, x_m) \rightarrow 0$ for $n, m \rightarrow \infty$.

The theorem below indicates that in any metric space E a convergent sequence is necessarily fundamental.

Theorem III.4.1. *If a sequence $\{x_n\}$ ($x_n \in E$; $i = 1, 2, \dots$) is convergent to a limit it is fundamental.*

Proof. Let $x_n \rightarrow x$. Then

$$\rho(x_n, x_m) \leq \rho(x_n, x) + \rho(x, x_m) \xrightarrow{n, m \rightarrow \infty} 0$$

Definition. A set of points of a metric space is said to be *bounded* if it is contained inside a sphere. In Sec. II.2 we state another definition of the boundedness of a set (for the space $\mathbf{R}_n$) equivalent to the definition stated here.

Theorem III.4.2. *Every fundamental sequence $\{x_n\}$ is bounded.*

Proof. Let us set an arbitrary $\varepsilon > 0$ and choose a number N such that $\rho(x_n, x_N) < \varepsilon$ for $n \geq N$. Now, putting $r = \max[\varepsilon, \rho(x_1, x_N), \rho(x_2, x_N), \dots, \rho(x_{N-1}, x_N)]$

we see that $\rho(x_n, x_N) \leq r$ for all $n = 1, 2, \dots$, i.e. for all $x_n \in S^*(x_N, r)$.*

Now we proceed to a definition playing an extremely important role in the theory of metric spaces.

Definition. A metric space is called *complete* if in this space every fundamental (Cauchy) sequence has a limit.

We have already shown that the space $\mathbf{R}_n$ is complete. Most of the spaces considered in the example in Sec. III.1 are also complete. As an instance, let us prove that the space $\mathcal{I}$ is complete.

Let $x_n = \{\xi_i^{(n)}\}$ and let $\{x_n\} \in \mathcal{I}$ be a fundamental sequence in the space $\mathcal{I}$. Since we have

$$|\xi_i^{(n)} - \xi_i^{(m)}| \leq \rho(x_n, x_m)$$

for each i we can also write $\xi_i^{(n)} - \xi_i^{(m)} \xrightarrow{n, m \rightarrow \infty} 0$ whence it follows that for each i there exists a finite $\xi_i = \lim \xi_i^{(n)}$. Now we must show that the sequence $x = \{\xi_i\}$ belongs to $\mathcal{I}$ and that $x = \lim x_n$.

By the hypothesis, for any $\varepsilon > 0$ there is N such that

$$\sum_{i=1}^{\infty} |\xi_i^{(n)} - \xi_i^{(m)}| < \frac{\varepsilon}{2}$$

for $n, m \geq N$. Retaining in this sum only a finite number p of terms we see that, given an arbitrary number p , there holds

$$\sum_{i=1}^p |\xi_i^{(n)} - \xi_i^{(m)}| < \frac{\varepsilon}{2}$$

for $n, m \geq N$. On fixing n and passing to the limit for $m \rightarrow \infty$ we receive $\sum_{i=1}^p |\xi_i^{(n)} - \xi_i| \leq \frac{\varepsilon}{2}$. This inequality

holding for any p , the series $\sum_{i=1}^{\infty} |\xi_i^{(n)} - \xi_i|$ is convergent

and $\sum_{i=1}^{\infty} |\xi_i^{(n)} - \xi_i| \leq \frac{\varepsilon}{2} < \varepsilon$. We have thus proved that $\{\xi_i^{(n)} - \xi_i\} \in \mathcal{I}$. It readily follows that $x = \{\xi_i\}$ also belongs

* Replacing r by $r' > r$ we see that all x_n are also contained in the open ball $S(x_N, r')$.

to l , and the inequality $\sum_{i=1}^{\infty} |\xi_i^{(n)} - \xi_i| < \varepsilon$ means that $\rho(x_n, x) < \varepsilon$ for $n \geq N$, which is what we set out to prove.

The completeness of the space l^2 is proved similarly.

Let us check the completeness of the space C . Let x_n ($n = 1, 2, \dots$) be a countable number of functions belonging to C and let $\rho(x_n, x_m) \rightarrow 0$. Then it is clear that the number sequence $\{x_n(t)\}$ is fundamental for every fixed value of t . Consequently there exists a limit $x(t) = \lim x_n(t)$.

Now, given an arbitrary $\varepsilon > 0$, we find N such that $\rho(x_n, x_m) < \varepsilon$ for $n, m \geq N$, i.e. $|x_n(t) - x_m(t)| < \varepsilon$ for all t and for all $n, m \geq N$. On passing to the limit in the last inequality for $m \rightarrow \infty$ we obtain $|x_n(t) - x(t)| \leq \varepsilon$ for all t and $n \geq N$, which means that $\{x_n\}$ converges to x uniformly. Therefore x belongs to C and $\rho(x_n, x) \rightarrow 0$.

By analogy with the above proof it can easily be shown that the space m is also complete.

As will be shown in Sec. VIII.5, the space C_L is not complete.

Every subset E_1 of a metric space E can be regarded as a new metric space if for any two points of E_1 we retain the definition of the distance introduced in E . In this connection we give the following theorem needed for our further aims.

Theorem III.4.3. *Every closed subset F of a complete metric space E is itself a complete metric space.*

Proof. Let $x_n \in F$ ($n = 1, 2, \dots$) and let $\{x_n\}$ be a fundamental sequence. The space E being complete, there exists $x = \lim x_n$ in E . Further, since F is closed there must be $x \in F$. Consequently, every fundamental sequence of points belonging to F has a limit in F , which means that the space F is complete.

Let us establish the following lemma which will be often used in what follows.

Lemma III.4.1. *If a fundamental sequence $\{x_n\}$ contains a subsequence $\{x_{n_i}\}$ convergent to an element x the whole sequence $\{x_n\}$ is also convergent to the same limit: $x_n \rightarrow x$.*

Proof. Taking some $n_i \geq n^*$ for any n we can write, on the basis of the triangle inequality,

$$\rho(x_n, x) \leq \rho(x_n, x_{n_i}) + \rho(x_{n_i}, x)$$

If $n \rightarrow \infty$ then also $n_i \rightarrow \infty$, and the right-hand side of the last inequality tends to zero. Consequently, we also have $\rho(x_n, x) \rightarrow 0$, that is $x_n \rightarrow x$.

§ 5. Separable Spaces

Consider the space $\mathbf{R}_1$ of all real numbers. The subset D of this space consisting of all rational numbers possesses the following property: every element of $\mathbf{R}_1$, that is every real number, is representable as the limit of a sequence of rational numbers, i.e. as the limit of a sequence of elements of the space $\mathbf{R}_1$ belonging to D . Thus, the closure of D coincides with $\mathbf{R}_1$: $\bar{D} = \mathbf{R}_1$. This property is described by saying that the set of rational numbers is *dense* in $\mathbf{R}_1$. Let us extend the notion of a dense set to arbitrary metric spaces.

Definition. A set A of points of a metric space E is said to be *everywhere dense* in E if $\bar{A} = E$.

As follows from Sec. III.3, the equality $\bar{A} = E$ means that, given any $x \in E$ and any $\varepsilon > 0$, there exists $x' \in A$ such that $\rho(x', x) < \varepsilon$. In its turn, this is equivalent to the property that every $x \in E$ is representable in the form $x = \lim x_n$ where $x_n \in A$.

Definition. A metric space is said to be *separable* if it contains a countable or finite subset everywhere dense in this space*.

Most of the spaces we dealt with in the foregoing examples were separable.

Every point in $\mathbf{R}_n$ is the limit of a sequence of points with rational coordinates. The points with rational coordinates form a countable set (see Theorem I.4.9, Sec. II.1) and consequently the space $\mathbf{R}_n$ is separable.

The separability of the space l is shown as follows. Let $x = \{\xi_i\} \in l$. Putting

$$x_n = \{\xi_1, \xi_2, \dots, \xi_n, 0, 0, \dots\}$$

we see that, according to the definition of convergence in l the sequence x_n is convergent to x . Furthermore, for any n there is a point $x'_n = (r_1, r_2, \dots, r_n, 0, 0, \dots)$ with rational coordinates lying arbitrarily close to x_n (in the

* It is readily seen that if the given space consists of an *infinite* number of points none of its everywhere dense subsets can be finite.

sense of the metric of l), for instance, such that $\rho(x'_n, x_n) < 1/n$. Then we also have $x'_n \rightarrow x$.

The set D_n of all the points of the form $(r_1, r_2, \dots, r_n, 0, 0, \dots)$ with rational coordinates is countable for any fixed n (see Theorem I.4.9). The union $\bigcup_{n=1}^{\infty} D_n$ is also countable (Theorem I.4.4) and, as has already been shown, it is everywhere dense in l . Consequently, the space l is separable.

The same argument can be applied to show that the space l^2 is also separable. As for the space m , it can be proved that it is nonseparable.

The space C is separable. As a countable and everywhere dense subset in this space we can take, for instance, the set $\mathcal{P}$ of all algebraic polynomials with rational coefficients.

Indeed, by the well-known *Weierstrass' polynomial approximation theorem* proved in courses of mathematical analysis* every function x continuous on an interval can be represented as the limit of a uniformly convergent sequence of polynomials Q_n . The uniform convergence of a functional sequence is nothing but the convergence with respect to the metric of the space C , that is $Q_n \rightarrow x$. On replacing all the coefficients of each of the polynomials Q_n by sufficiently close rational numbers we obtain some polynomials $P_n(t)$ belonging to the set $\mathcal{P}$ (see Sec. I.4) such that

$$|P_n(t) - Q_n(t)| < \frac{1}{n}$$

throughout the interval $a \leq t \leq b$. Thus, $\rho(P_n, Q_n) < \frac{1}{n}$ whence it follows that $P_n \rightarrow x$ in C . We have thus shown that the countable set $\mathcal{P}$ is everywhere dense in C .

Since the convergence in C implies the convergence in C_L (see Sec. III.2) the set $\mathcal{P}$ is also everywhere dense in C_L , and hence the space C_L is separable.

* For instance, see B. M. Budak, S. V. Fomin, *Multiple Integrals, Field Theory and Series (An Advanced Course of Higher Mathematics)*, Moscow, Mir Publishers, 1973 (English edition) and W. Rudin, *Principles of Mathematical Analysis*, New York, McGraw-Hill, 1964.

§ 6. Normed Spaces

The concept of a *normed space* is one of the most important notions of functional analysis. The development of the theory of normed spaces is mostly due to the works of S. Banach* published in the twenties. In the present course we shall confine ourselves to normed *function spaces*, that is to spaces whose elements are functions, since only this class of spaces is needed for our aims. Functional analysis also deals with more general normed spaces of abstract elements.

Let T be an arbitrary set and $\mathbf{X}$ a nonempty family of real functions $x(t)$ defined on T and assuming finite values. We say that the set $\mathbf{X}$ is a *linear (function) space* if it is closed with respect to ordinary operations of addition of functions and multiplication of functions by numbers.** In short, this means that, given any two functions x and y belonging to the set $\mathbf{X}$ and an arbitrary real number λ , the functions $x + y$ and λx also belong to $\mathbf{X}$.

If we take as T the set of all natural numbers n the functions $x(t)$ turn into infinite number sequences $x_n = x(n)$; if T is the set of the first n natural numbers the functions on T are the points of the n -dimensional space of n -tuples. Therefore all that is said below in connection with general function spaces is also applicable to the sets of sequences and to the sets of points of the n -dimensional space.

The definition readily shows that if $\mathbf{X}$ is a linear (function) space then for any finite number of functions $x_1, x_2, \dots, x_k$ belonging to $\mathbf{X}$ and any real numbers $\lambda_1, \lambda_2, \dots, \lambda_k$ the

function $x(t) = \sum_{i=1}^k \lambda_i x_i(t)$ (for brevity also written $x = \sum_{i=1}^k \lambda_i x_i$) belongs to $\mathbf{X}$. In particular, the difference of

* S. Banach (1892-1945), a famous Polish mathematician, one of the founders of functional analysis.

** Functional analysis also deals with a more general concept of a linear space composed of arbitrary elements for which the arithmetical operations are defined axiomatically. E.g., see B. Z. Vulikh, *Introduction to Functional Analysis*, Moscow, "Nauka", 1967 (Russian edition), § 5.1, and K. Yosida, *Functional Analysis*, Springer, 1966 (English edition).

any two functions belonging to $\mathbf{X}$ also belongs to $\mathbf{X}$. The function identically equal to zero on T can be represented as the product of any function belonging to $\mathbf{X}$ by the number 0 and therefore must be contained in $\mathbf{X}$. This function will be referred to as the *zero element* of $\mathbf{X}$ and denoted θ .

There are many linear function spaces in which it is possible to introduce a metric in some special manner so that the basic role is played not by the notion of the distance function but by a new notion of the *norm* of an element.

Definition. A linear (function) space $\mathbf{X}$ is called a *normed space* if to each element $x \in \mathbf{X}$ there is assigned a definite real number called the *norm* of x and denoted $\|x\|$ so that the following conditions (*axioms of norm* or *axioms of normed space*) hold:

- I. $\|x\| \geq 0$ for any $x \in \mathbf{X}$; $\|x\| = 0$ only if $x = \theta$;
- II. $\|\lambda x\| = |\lambda| \|x\|$ for any $x \in \mathbf{X}$ and any number λ ;
- III. $\|x + y\| \leq \|x\| + \|y\|$ for any $x, y \in \mathbf{X}$ (this condition is also called the *triangle inequality*).

Now, given a normed space $\mathbf{X}$, we can introduce a metric in it by putting

$$\rho(x, y) = \|x - y\| \quad (7)$$

for any two elements $x, y \in \mathbf{X}$. This definition shows that, since $x - \theta = x$, we have

$$\|x\| = \|x - \theta\| = \rho(x, \theta)$$

that is *the norm of any element is equal to its distance from θ* .

Let us show that $\mathbf{X}$ satisfies all the axioms of metric space.

If $x \neq y$ then $x - y \neq \theta$, and formula (7) and condition I of this section show that $\rho(x, y) > 0$; if $x = y$ we have $x - y = \theta$, and hence $\rho(x, y) = 0$. Furthermore, we have $x - y = (-1)(y - x)$, and therefore, by Axiom II, $\|x - y\| = \|y - x\|$, that is $\rho(x, y) = \rho(y, x)$. Finally, the triangle inequality for the distance follows from that for the norm:

$$\begin{aligned} \rho(x, y) &= \|x - y\| = \|(x - z) + (z - y)\| \leq \\ &\leq \|x - z\| + \|z - y\| = \rho(x, z) + \rho(z, y) \end{aligned}$$

Thus, a normed space is a special case of a metric space.

As an example of a normed space we can take the space R_n if we put $\|x\| = \sqrt{\sum_{i=1}^n \xi_i^2}$ for each $x = (\xi_1, \xi_2, \dots, \xi_n)$.

The first two axioms of norm obviously hold in this case (the zero element of R_n is the point θ whose all coordinates are equal to 0) while the triangle inequality for norm takes the form of the Cauchy inequality (see inequality (4) proved in Sec. II.1). If, proceeding from the norm defined above, we introduce the distance in R_n by means of formula (7) this metric will coincide with that defined directly for R_n in Sec. II.1.

Another example of a normed space is the space C of continuous functions on an interval $[a, b]$ if we put $\|x\| = \max |x(t)|$.

In this case the fulfilment of the first two requirements of the definition of norm is also obvious while the triangle inequality can be verified by analogy with Sec. III.1 where it was established for the distance function in the space C . Formula (7) leads to the same distance function in C as the one introduced for C in Sec. III.1.

Later on we shall encounter some other normed function spaces.

Applying to a normed space the general definition of convergence stated for a metric space we conclude that

$$x_m \rightarrow x \text{ means that } \|x_m - x\| \rightarrow 0$$

This type of convergence is referred to as *convergence in norm*. The general properties of the distance function in a metric space show that the function $\|x - y\|$ is continuous with respect to its arguments x and y , that is if $x_m \rightarrow x$, $y_m \rightarrow y$ then $\|x_m - y_m\| \rightarrow \|x - y\|$ (see Sec. III.2). In particular, putting $y_m = \theta$ for all m we see that $\|x_m\| \rightarrow \|x\|$ for $x_m \rightarrow x$ (the *continuity of the norm*).

The propositions below express the continuity property of the basic arithmetical operations in a normed space.

(a) If $x_m \rightarrow x$ and $y_m \rightarrow y$ then $x_m + y_m \rightarrow x + y$.
Indeed,

$$\begin{aligned} \|(x_m + y_m) - (x + y)\| &= \|(x_m - x) + (y_m - y)\| \leq \\ &\leq \|x_m - x\| + \|y_m - y\| \rightarrow 0 \end{aligned}$$

(b) If $x_m \rightarrow x$ and $\lambda_m \rightarrow \lambda$ then $\lambda_m x_m \rightarrow \lambda x$.

Indeed

$$\begin{aligned} \|\lambda_m x_m - \lambda x\| &= \|\lambda_m (x_m - x) + (\lambda_m - \lambda) x\| \leq \\ &\leq |\lambda_m| \|x_m - x\| + |\lambda_m - \lambda| \|x\| \rightarrow 0 \end{aligned}$$

For a bounded set in a normed space X we can restate the general definition, given in Sec. III.4 for arbitrary metric spaces, in a simpler form: *a set $A \subset X$ is bounded if the norms of all its elements are uniformly bounded* (i.e. by one and the same positive number). Indeed, if A is bounded, then, in accordance with the general definition, it is contained in a sphere: $A \subset S^*(x_0, r)$. Consequently, $\|x - x_0\| \leq r$ for all $x \in A$, and therefore

$$\begin{aligned} \|x\| &= \|(x - x_0) + x_0\| \leq \|x - x_0\| + \|x_0\| \leq \\ &\leq r + \|x_0\| \end{aligned}$$

Conversely, if $\|x\| \leq R$ for all $x \in A$ we have $A \subset S^*(0, R)$, and thus A is bounded according to the general definition.

Following the definition of the distance in a normed space we can characterize a fundamental sequence $\{x_m\}$ belonging to a normed space by the condition

$$\|x_m - x_p\| \rightarrow 0 \quad \text{for } m, p \rightarrow \infty$$

Theorem III.4.2 implies that *the norms of the elements of any fundamental sequence are uniformly bounded*.

A complete normed space, that is one in which every fundamental sequence is convergent to a limit, is called a *Banach space*.

For an arbitrary normed space, we can consider series composed of elements of the space: $\sum_{m=1}^{\infty} x_m$. The sum of such a series is defined in the ordinary way as the limit of the sequence of its partial sums; if this limit exists the series is said to be *convergent*.

§ 7. Linear Functionals in Normed Spaces

Generalizing the notion of a function of one independent variable we state the following definition: a mapping of a metric space (or of its part) into the set of real or complex

numbers is called a *functional**. In our course we shall only deal with functionals taking real values and therefore this assumption will not be stipulated in what follows.

As an example of a functional defined on the space C we can take the integral $\int_a^b x(t) dt$. The value $x(t_0)$ of a function $x(t)$ at a fixed point t_0 is also a functional on C .

We sometimes also deal with functionals dependent on several arguments. For instance, such is the distance function $\rho(x, x')$ in a metric space X which can be interpreted as a functional, dependent on two arguments x and x' , defined for any pair of points belonging to the space X .

The definition of continuity stated in Sec. II.7 for functions in a Euclidean space can be extended to functionals on any metric spaces.

Definition. A functional f defined on a set $D \subset X$ is said to be *continuous at a point* $x_0 \in D$ if for any sequence of points $x_m \in D$ such that $x_m \rightarrow x_0$ there holds the relation $f(x_m) \rightarrow f(x_0)$. A functional is simply called *continuous* if it is continuous at every point of the set it is defined on.

This definition of continuity of a functional has been stated in terms of sequences. It can also be restated in "ε, δ language": a functional f is *continuous at a point* $x_0 \in D$ if, given any $\varepsilon > 0$, there is $\delta > 0$ such that the inequality $|f(x) - f(x_0)| < \varepsilon$ holds for all $x \in D$ satisfying the condition $\rho(x, x_0) < \delta$. The equivalence of the latter definition to the former (stated in terms of sequences) is established by means of the standard procedure used in classical mathematical analysis.

The norm of the element of a normed function space is a continuous functional; this was in fact proved in Sec. III.6.

In the theory of normed spaces an extremely important role is played by the so-called *linear functionals*. The considerations below are based on the assumption that the normed space X in question contains elements distinct from the zero element.**

* By the way, such a mapping is also called a *function*.

** The set X consisting of the single function $\theta(t) \equiv 0$ for which we put $\|\theta\| = 0$ is obviously a normed space. In our discussion in the present section we exclude this case.

Definition. A functional f defined on a normed space X is called *linear* if it possesses the following properties:

(a) *distributivity*: for any $x_1, x_2, \dots, x_h \in X$ and any numbers $\lambda_1, \lambda_2, \dots, \lambda_h$ the relation

$$f\left(\sum_{i=1}^h \lambda_i x_i\right) = \sum_{i=1}^h \lambda_i f(x_i)$$

holds;

(b) *continuity* (understood in the sense of the above definitions).

A functional f is said to be *additive* if it satisfies the condition

$$f(x_1 + x_2) = f(x_1) + f(x_2) \quad \text{for any } x_1, x_2 \in X$$

It can be proved that if an additive functional is continuous it is *homogeneous* in the sense that

$$f(\lambda x) = \lambda f(x) \quad \text{for any } x \in X \text{ and any number } \lambda$$

The last assertion shows that a continuous additive functional is distributive but in what follows we shall not use this fact since in the cases we shall deal with the distributivity of the functionals in question can easily be checked.

It should be noted that for every distributive functional and even for every additive functional we have $f(\theta) = 0$. Indeed, taking any $x \in X$ we can write

$$f(x) = f(x + \theta) = f(x) + f(\theta)$$

whence follows $f(\theta) = 0$.

Theorem III.7.1. *If a distributive functional f is continuous at a point $x_0 \in X$ it is also continuous at any other point $x \in X$, that is, it is linear.**

Proof. Let $x_m \in X$ and $x_m \rightarrow x$. Then, by the continuity of the arithmetical operations in X , we have $x_m - x + x_0 \rightarrow x_0$, and the continuity of the functional at the point x_0 therefore indicates that

$$f(x_m) - f(x) + f(x_0) \rightarrow f(x_0)$$

whence $f(x_m) \rightarrow f(x)$.

* It can easily be shown that in this theorem the condition of distributivity of the functional can be replaced by the weaker condition of additivity.

This theorem allows us to confine ourselves to the verification of the continuity of a given distributive functional at a single point, for instance, at the point θ when it is necessary to check its linearity. The continuity at the point θ simply means that $f(x_m) \rightarrow 0$ if $\|x_m\| \rightarrow 0$.

Let us prove that *in the Euclidean space $\mathbf{R}_n$ every distributive functional is linear*. Indeed, the addition of two elements of $\mathbf{R}_n$ is equivalent to the addition of their similar coordinates, that is if $x = (\xi_1, \xi_2, \dots, \xi_n)$ and $y = (\eta_1, \eta_2, \dots, \eta_n)$ then

$$x + y = (\xi_1 + \eta_1, \xi_2 + \eta_2, \dots, \xi_n + \eta_n)$$

Analogously,

$$\lambda x = (\lambda \xi_1, \lambda \xi_2, \dots, \lambda \xi_n)$$

If we introduce the "unit vectors e_i along the coordinate axes" in $\mathbf{R}_n$ ($i = 1, 2, \dots, n$) which are the elements of $\mathbf{R}_n$ whose i th ($i = 1, 2, \dots, n$) coordinate is equal to 1 while the other coordinates are equal to 0, every element $x = (\xi_1, \xi_2, \dots, \xi_n) \in \mathbf{R}_n$ can be represented in the form

$$x = \sum_{i=1}^n \xi_i e_i$$

Let f be a distributive functional in $\mathbf{R}_n$. Let us put $f(e_i) = \alpha_i$ ($i = 1, 2, \dots, n$). Then

$$f(x) = \sum_{i=1}^n \xi_i f(e_i) = \sum_{i=1}^n \alpha_i \xi_i$$

for any $x \in \mathbf{R}_n$. Let us take a sequence $x_m = (\xi_1^{(m)}, \xi_2^{(m)}, \dots, \xi_n^{(m)}) \rightarrow \theta$. This means that $\xi_i^{(m)} \xrightarrow{m \rightarrow \infty} 0$ for each i . Then we also have

$$f(x_m) = \sum_{i=1}^n \alpha_i \xi_i^{(m)} \rightarrow 0$$

i.e. the functional f is continuous at the zero point θ and thus, by virtue of the foregoing theorem, it is linear.

It is clear that for any $\alpha_1, \alpha_2, \dots, \alpha_n$ the formula

$$f(x) = \sum_{i=1}^n \alpha_i \xi_i$$

determines a distributive functional in R_n . Thus, this formula gives the general representation of a linear functional in R_n .

In general normed spaces the property of continuity of a distributive functional turns out to be equivalent to another property, referred to as *boundedness* which is defined as follows.

Definition. A functional f defined on a normed space X is said to be *bounded* if there is a nonnegative constant C such that

$$|f(x)| \leq C \|x\| \quad (8)$$

for all $x \in X$.

Theorem III.7.2. A distributive functional f is linear if and only if it is bounded.*

Proof. (a) *Necessity.* Let the functional f be linear. Suppose that it is not bounded. Then, for any natural number m , there exists $x_m \in X$ such that

$$f(x_m) > m \|x_m\|$$

This inequality shows that $f(x_m) \neq 0$, and hence $x_m \neq \theta$. Let us put $x'_m = \frac{x_m}{m \|x_m\|}$. Then $\|x'_m\| = \frac{1}{m} \rightarrow 0$, i.e. $x'_m \rightarrow \theta$; at the same time

$$f(x'_m) = \frac{f(x_m)}{m \|x_m\|} > 1$$

which contradicts the continuity of the functional f . Thus, the functional f is bounded.

(b) *Sufficiency.* Let f be a bounded distributive functional. This means that condition (8) holds. If $x_m \rightarrow \theta$, i.e. $\|x_m\| \rightarrow 0$, condition (8) readily implies that we must also have $f(x_m) \rightarrow 0$. Hence, the functional f is continuous at the point θ and therefore, by virtue of Theorem III.7.1, it is linear.

Definition. Given a linear functional f , the least of the constants C for which condition (8) is fulfilled for all $x \in X$ is called the *norm* of the functional f and is denoted by $\|f\|$.

Let us prove that the least constant C exists and is equal to

$$\sup_{x \in S^*} |f(x)|$$

* This theorem applies to any additive functional as well.

where S^* is the *unit sphere* in the space X , that is the set of all $x \in X$ with $\|x\| \leq 1$. In other words, $\|f\|$ is equal to the least upper bound of the set of values of $|f(x)|$ on the sphere S^* .

Indeed, if C is a constant satisfying condition (8) we have

$$|f(x)| \leq C \|x\| \leq C$$

for any $x \in S^*$, and therefore

$$C_1 = \sup_{x \in S^*} |f(x)| \leq C$$

Let us put $x' = \frac{x}{\|x\|}$ for any $x \neq \theta$. Then $x' \in S^*$, and consequently $|f(x')| \leq C_1$ and

$$|f(x)| = |f(\|x\| x')| = \|x\| |f(x')| \leq C_1 \|x\|$$

Thus, for the constant C_1 as well condition (8) is fulfilled for all $x \neq \theta$. For $x = \theta$ this condition is trivial. Consequently C_1 is nothing but the least value of C in condition (8), that is $C_1 = \|f\|$.

From the definition of the norm it follows, in particular, that

$$|f(x)| \leq \|f\| \|x\|$$

for any $x \in X$.

As an example of a linear functional in the space C we can take $f(x) = x(t_0)$ where t_0 is any fixed point belonging to the domain of definition of x . For this functional we have $\|f\| = 1$. If C consists of all continuous functions on an interval $[a, b] \subset R_1$ and $f(x) = \int_a^b x dt$ we readily conclude that f is a linear functional and $\|f\| = b - a$.

If a series $\sum_{m=1}^{\infty} x_m$ formed of elements of a normed space X is convergent and x is its sum, that is $x = \sum_{m=1}^{\infty} x_m$, the relation

$$f(x) = \sum_{m=1}^{\infty} f(x_m) \quad (9)$$

holds for any linear functional f on X . Indeed, $x = \lim \sigma_m$ where $\sigma_m = \sum_{i=1}^m x_i$, and thus

$$f(x) = \lim f(\sigma_m) = \lim \sum_{i=1}^m f(x_i) = \sum_{i=1}^{\infty} f(x_i)$$

Measure in Abstract Sets

§ 1. Additive Set Functions

The notion of a measure in a Euclidean space is a natural generalization of such elementary notions as the length of a line segment, the area of a rectangle and the volume of a parallelepiped. This general notion of a measure is introduced to construct a more general concept of integral than that of the Riemann integral studied in classical mathematical analysis. In this chapter we shall deal with a still more general concept of a measure in an arbitrary abstract set. An important special case of the general process of constructing a measure in an abstract set is the construction of the Lebesgue measure in a Euclidean space which will be considered in the next chapter.

A function whose domain of definition is a system of sets is referred to as a *set function*. Suppose that a real function f is defined for all sets A belonging to a system of sets $\mathfrak{R}$. In the general case we allow $f(A)$ to take on the values $+\infty$ and $-\infty$; to simplify the further definitions we shall agree that if f assumes infinite values they are all of one definite sign, for instance $+\infty$. A function f of this kind is said to be σ -*additive* (also *countably additive* or *completely additive*) if for any finite or countable family of disjoint sets $A_i \in \mathfrak{R}$ whose union $A = \bigcup_i A_i$ also belongs to $\mathfrak{R}$ there holds the equality

$$f(A) = \sum_i f(A_i) \quad (1)$$

If equality (1) is only guaranteed when A is a finite union of disjoint sets A_i (all A_i and A belong to $\mathfrak{R}$) the function f is called *finitely additive* or, simply, *additive*.

* When any finite number or $+\infty$ is added to $+\infty$ the resultant sum is naturally considered equal to $+\infty$.

Let the domain of definition $\mathfrak{R}$ of a set function f be a ring of sets. Then the union of any finite number of sets $A_i \in \mathfrak{R}$ also belongs to $\mathfrak{R}$. Therefore, in this case we conclude that if the relation

$$f(A_1 \cup A_2) = f(A_1) + f(A_2)$$

holds for any two disjoint sets A_1 and A_2 belonging to $\mathfrak{R}$ then, by induction, equality (1) readily follows for any finite number of disjoint sets $A_i \in \mathfrak{R}$, and the function f is thus finitely additive.

Additive functions defined on a ring of sets possess the following property: if $A, B \in \mathfrak{R}$, $B \subset A$ and the value $f(A)$ is finite then $f(B)$ is finite as well.

Indeed, the additivity of the function f implies that

$$f(A) = f(B) + f(A \setminus B)$$

(the set $A \setminus B \in \mathfrak{R}$ since $\mathfrak{R}$ is a ring). If we assume that $f(B) = +\infty$ then $f(A)$ is also equal to $+\infty$, which contradicts the hypothesis.

Let $A, B \in \mathfrak{R}$, $B \subset A$, and let us suppose that $f(B)$ is finite. Then the foregoing formula readily shows that

$$f(A \setminus B) = f(A) - f(B)^* \quad (2)$$

Putting $B = A$ in formula (2) we obtain

$$f(\emptyset) = f(A) - f(A) = 0$$

Thus, the equality $f(\emptyset) = 0$ necessarily holds if there is at least one set $A \in \mathfrak{R}$ for which $f(A)$ is finite.

σ -additive set functions with finite values can be characterized by means of the following theorem.

Theorem IV.1.1. *For an additive set function f with finite values defined on a ring $\mathfrak{R}$ to be σ -additive it is necessary and sufficient that for any decreasing sequence of sets $A_i \in \mathfrak{R}$ ($i = 1, 2, \dots$) whose intersection is empty (i.e. $\bigcap_{i=1}^{\infty} A_i = \emptyset$) the limiting relation*

$$f(A_i) \rightarrow 0$$

hold.

* If we had $f(A) = f(B) = +\infty$ formula (2) would make no sense.

Proof. If f is σ -additive then from formula (3) of Chapter I and also from formula (2) established above it follows that for any decreasing sequence of sets $A_i \in \mathfrak{R}$ with empty intersection there must be

$$\begin{aligned} f(A_1) &= \sum_{i=1}^{\infty} [f(A_i) - f(A_{i+1})] = \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^{n-1} [f(A_i) - f(A_{i+1})] = f(A_1) - \lim_{n \rightarrow \infty} f(A_n) \end{aligned}$$

Consequently, $f(A_n) \rightarrow 0$.

Conversely, let the condition of the theorem be fulfilled and let $A \in \mathfrak{R}$, $A = \bigcup_{i=1}^{\infty} A_i$ where the sets $A_i \in \mathfrak{R}$ are disjoint. Putting $B_n = \bigcup_{i=n+1}^{\infty} A_i$ we can write $\bigcap_{n=1}^{\infty} B_n = \emptyset$ and

$$B_n = A \setminus (A_1 \cup A_2 \cup \dots \cup A_n) \in \mathfrak{R}$$

and the finite additivity of the function f implies that

$$f(A) = \sum_{i=1}^n f(A_i) + f(B_n)$$

By the hypothesis, we have $f(B_n) \rightarrow 0$ (see proposition 2° in Sec. I.2), and therefore $f(A) = \sum_{i=1}^{\infty} f(A_i)$. Hence, the function f is σ -additive.

Theorem IV.1.2. Let f be a σ -additive set function defined on a ring $\mathfrak{R}$. If $A \in \mathfrak{R}$ and $A = \bigcup_{i=1}^{\infty} A_i$ where the sets A_i belong to $\mathfrak{R}$ and form an increasing sequence, the limiting relation

$$f(A) = \lim_{i \rightarrow \infty} f(A_i) \quad (3)$$

must hold. The same equality (3) is fulfilled if $A = \bigcap_{i=1}^{\infty} A_i$ ($A, A_i \in \mathfrak{R}$), the sets A_i form a decreasing sequence and $f(A_i)$ is finite beginning with some i .

Proof. Let us consider the case of an increasing sequence and begin with the case when all the values $f(A_i)$ are finite. Then it follows from formula (5) of Chapter I that

$$f(A) = f(A_1) + \sum_{i=2}^{\infty} [f(A_i) - f(A_{i-1})]$$

which is equivalent to (3). Now, if $f(A_i) = +\infty$ beginning with some i then $f(A) = +\infty$ and equality (3) is obviously true.

In the case of a decreasing sequence we can assume, without loss of generality, that $f(A_1)$ is finite. Then Theorem IV.1.1 can be applied to the restriction of the function f to the part of the ring $\mathfrak{R}$ which consists of the subsets contained in A_1 . Since we obviously have $\bigcap_{i=1}^{\infty} (A_i \setminus A) = \emptyset$, this theorem shows that $f(A_i \setminus A) \rightarrow 0$, which again is equivalent to (3).

§ 2. Definition of a Measure. Properties of Measures

We begin this section with a definition playing a basic role in our further investigations.

Definition. Let X be an arbitrary abstract set. By a *measure* in X we mean a real nonnegative σ -additive function m , defined on a semiring $\mathfrak{R}$ of subsets of the set X , such that $m\emptyset = 0$.

A measure m is said to be *finite* if $mA < +\infty$ for any $A \in \mathfrak{R}$. It is called *σ -finite* if for any $A \in \mathfrak{R}$ there exists a system of subsets $A_n \in \mathfrak{R}$ ($n = 1, 2, \dots$) such that $A \subset \bigcup_n A_n$ and $mA_n < +\infty$ for every n .

Note that for a σ -finite measure the condition $m\emptyset = 0$ can be deduced from the other properties of measure. But this is not so in the general case: a nonnegative σ -additive function can be identically equal to $+\infty$.

Let us enumerate some properties of a measure which readily follow from its definition.

(a) If we are given an at most countable family of disjoint sets $A_n \in \mathfrak{R}$ ($n = 1, 2, \dots$)* such that $A_n \subset A \in \mathfrak{R}$ for any n then $\sum_n mA_n < mA$.

Indeed, if the number of the sets A_n is finite (for definiteness, let this number be p) then, according to proposition (a) in Sec. I.7, the difference $A \setminus \bigcup_{n=1}^p A_n$ can be represented in the form

$$A \setminus \bigcup_{n=1}^p A_n = \bigcup_k C_k$$

where the sets $C_k \in \mathfrak{R}$ are disjoint. It follows that

$$mA = \sum_{n=1}^p mA_n + \sum_k mC_k$$

and consequently

$$\sum_{n=1}^p mA_n \leq mA$$

If the family of the sets A_n is countable the above inequality holds for any p ; passing to the limit for $p \rightarrow \infty$ we obtain

$$\sum_{n=1}^{\infty} mA_n \leq mA$$

A special case of property (a) is the *monotonicity* of measure: if $A, B \in \mathfrak{R}$ and $B \subset A$ then $mB \leq mA$.

(b) If $A, A_n \in \mathfrak{R}$ ($n = 1, 2, \dots$), $A \subset \bigcup_n A_n$ and the system of the sets A_n is finite or countable then

$$mA \leq \sum_n mA_n$$

To prove this property let us put $B_n = A_n \cap A$; then $B_n \in \mathfrak{R}$ and $A = \bigcup_n B_n$. We also introduce the sets

$$D_1 = B_1, \quad D_2 = B_2 \setminus B_1, \quad D_3 = B_3 \setminus (B_1 \cup B_2), \dots$$

* Since we speak of an at most countable family of sets A_n the expression " $n = 1, 2, \dots$ " should be understood as indicating that n assumes natural values beginning with unity, which does not exclude the possibility that there are only a finite number of different values of n , for instance, $n = 1, 2, \dots, p$.

According to proposition 3° in Sec. I.2, we have $A = \bigcup_n D_n$, and the sets D_n are disjoint. Furthermore, every D_n is representable as $D_n = \bigcup_k C_{nk}$ where $C_{nk} \in \mathfrak{R}$ ($k = 1, 2, \dots$) and are disjoint. Then $A = \bigcup_{n,k} C_{nk}$, the sets C_{nk} entering into this union being disjoint. Besides, $D_n \subset B_n$, and therefore, by the foregoing proposition (a), we have

$$\sum_k mC_{nk} \leq mB_n$$

for any n whence

$$mA = \sum_n \sum_k mC_{nk} \leq \sum_n mB_n \leq \sum_n mA_n$$

Property (b) of a measure proved above is referred to as *countable semiadditivity*. This property immediately implies that if a finite or countable union of sets of measure zero belongs to $\mathfrak{R}$ the measure of the union is also equal to zero.

§ 3. Outer Measure

In this section we shall study a set function playing an important role in the measure theory.

Definition. Let X be an arbitrary set. By an *outer measure* in X is meant a real nonnegative set function μ^* defined on the family of all subsets of the set X and satisfying the following conditions:

(i) $\mu^* \emptyset = 0$;

(ii) if $E \subset \bigcup_n E_n$ where the system of the sets $E_n \subset X$

is at most countable then $\mu^* E \leq \sum_n \mu^* E_n$ (*countable semiadditivity* of the outer measure).

In particular, the countable semiadditivity of the outer measure implies its *monotonicity*: if $E \subset E_1$ then $\mu^* E \leq \mu^* E_1$.

It should be stressed that on the outer measure* neither the condition of σ -additivity nor even the condition of finite additivity is imposed.

* A measure in X is an outer measure in the case when it is defined on the family of all subsets of X .

Now we are going to study two procedures which, respectively, make it possible to construct an outer measure from a given measure and a measure from a given outer measure.

Theorem IV.3.1. *Let m be a measure in X defined on a semiring $\mathfrak{R}$ and let μ^* be the set function defined for any $E \subset X$ according to the following rule:*

(i) *if there exists an at most countable covering of E by sets belonging to the semiring $\mathfrak{R}$, that is if $E \subset \bigcup_n A_n$ where $A_n \in \mathfrak{R}$ ($n = 1, 2, \dots$), then*

$$\mu^*E = \inf \left\{ \sum_n mA_n \right\}$$

where the infimum is taken over all the possible coverings of the indicated type;

(ii) *if otherwise, $\mu^*E = +\infty$.*

Then μ^ is an outer measure in X , and $\mu^*A = mA$ for any $A \in \mathfrak{R}$.*

Proof. The fulfilment, for μ^* , of the first condition of the definition of outer measure is obvious. The second condition of this definition only needs verification for the case when $\sum_n \mu^*E_n < +\infty$. In this case, given any $\varepsilon > 0$, for each E_n there exists a covering consisting of sets $A_{nh} \in \mathfrak{R}$ ($h = 1, 2, \dots$) such that

$$\sum_h mA_{nh} < \mu^*E_n + \frac{\varepsilon}{2^n}$$

The family of all sets A_{nh} ($n, h = 1, 2, \dots$) forms an at most countable covering of the set E (recall that $E \subset \bigcup_n E_n$ in the second condition of the definition of outer measure), and

$$\mu^*E \leq \sum_{n,h} mA_{nh} < \sum_n \left(\mu^*E_n + \frac{\varepsilon}{2^n} \right) \leq \sum_n \mu^*E_n + \varepsilon$$

Now the desired inequality follows from the arbitrariness of ε . Thus, μ^* is an outer measure.

The equality $\mu^*A = mA$ for $A \in \mathfrak{R}$ is obtained on the basis of the following considerations. On the one hand, by

what was proved in (b), Sec. IV.2, we have

$$mA \leq \sum_n mA_n$$

for any covering of the set A , and therefore $mA \leq \mu^*A$. On the other hand, the system of sets containing the set A alone is its own covering and therefore $\mu^*A \leq mA$. Hence, $\mu^*A = mA$. The theorem has been proved.

In what follows we shall say that the outer measure μ^* constructed in accordance with Theorem IV.3.1 is *generated by the measure m* .

Before proceeding to the construction of a measure from a given outer measure let us introduce the following notation

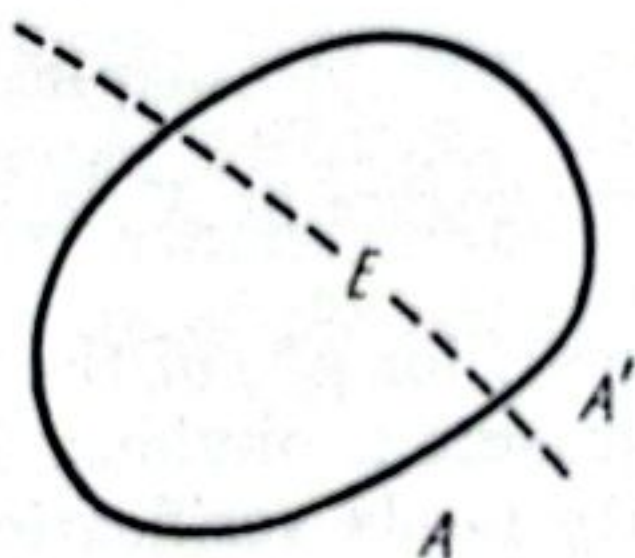


Fig. 8

which will be used *throughout the present chapter*: if E is an arbitrary subset of X then E' denotes the complement of E with respect to X : $E' = X \setminus E$.

Now, let μ^* be an arbitrary outer measure in X . Let us consider two sets $A, E \in X$. Suppose that A is such that the following characteristic property holds:

$$\mu^*E = \mu^*(E \cap A) + \mu^*(E \cap A') \quad (4)$$

(see Fig. 8). In this case we shall say that the set A “divides well” the set E . It should be noted that the property of countable semiadditivity of the outer measure indicates that we can always write the sign $\leq$ in formula (4). Therefore, to check whether A divides well E it suffices to establish the inequality

$$\mu^*E \geq \mu^*(E \cap A) + \mu^*(E \cap A') \quad (5)$$

which only needs verification in the case $\mu^*E < +\infty$.

HENRIQUE LAZARI

Since A and A' are involved symmetrically in formula (4), if, for the given E , the set A possesses property (4) then A' possesses it as well.

A set $A \subset X$ will be called μ^* -measurable if it divides well any set $E \subset X$. The restriction* of the outer measure μ^* to the family of all μ^* -measurable sets will be denoted by μ .

Theorem IV.3.2. *The system $\mathfrak{S}$ of all μ^* -measurable subsets of X is a σ -algebra, and the function μ is a measure in X .*

Proof. (a) We begin with proving that $\mathfrak{S}$ is an algebra and that the function μ is finitely additive.

It is clear that the set X itself and the empty set $\emptyset$ are contained in $\mathfrak{S}$, and hence the system of sets $\mathfrak{S}$ is nonempty. Besides, by what was said in connection with property (4), it readily follows that if $A \in \mathfrak{S}$ then $A' \in \mathfrak{S}$ as well.

Let $A_1, A_2 \in \mathfrak{S}$ and $B = A_1 \cap A_2$; let us show that B divides well any $E \subset X$. Using the fact that A_1 and A_2 divide well any set E we can write

$$\begin{aligned} \mu^* E &= \mu^* (E \cap A_1) + \mu^* (E \cap A_1') = \\ &= \mu^* (E \cap A_1 \cap A_2) + \mu^* (E \cap A_1 \cap A_2') + \mu^* (E \cap A_1') \end{aligned}$$

On the other hand,

$$\begin{aligned} \mu^* (E \cap B) + \mu^* (E \cap B') &= \\ &= \mu^* (E \cap B) + \mu^* (E \cap B' \cap A_1) + \mu^* (E \cap B' \cap A_1') \end{aligned}$$

But $B' \supset A_1'$ and $B' \cap A_1 = A_1 \cap A_2'$, and therefore the right-hand members of the last two equalities coincide; consequently,

$$\mu^* E = \mu^* (E \cap B) + \mu^* (E \cap B')$$

Thus, $B \in \mathfrak{S}$ and $\mathfrak{S}$ is an algebra (see the remark made after Theorem I.7.1).

* If F is an arbitrary mapping defined on a set $\mathfrak{M} = \{E\}$ and $\mathfrak{N} \subset \mathfrak{M}$, the restriction of the mapping F to the set $\mathfrak{N}$ is the mapping Φ , defined on $\mathfrak{N}$, which is determined by the formula $\Phi(E) = F(E)$ for $E \in \mathfrak{N}$ (this means that we consider the same mapping F , that is the values $\Phi(E)$ coincide with $F(E)$, but the domain of definition of Φ is narrower than that of F).

Finally, let $A_1, A_2 \in \mathfrak{S}$, $A_1 \cap A_2 = \emptyset$ and $A = A_1 \cup A_2$. Then, since the set A_1 is μ^* -measurable, we have

$$\mu^*(E \cap A) = \mu^*(E \cap A \cap A_1) + \mu^*(E \cap A \cap A_1')$$

for any $E \subset X$. At the same time, $A \cap A_1 = A_1$ and $A \cap A_1' = A_2$; consequently

$$\mu^*(E \cap A) = \mu^*(E \cap A_1) + \mu^*(E \cap A_2) \quad (6)$$

In particular, putting $E = A$ we find

$$\mu A = \mu^* A = \mu^* A_1 + \mu^* A_2 = \mu A_1 + \mu A_2$$

The finite additivity of the function μ has thus been proved.

(b) Now we shall show that $\mathfrak{S}$ is a σ -algebra and that the function μ is σ -additive.

First, let $A_k \in \mathfrak{S}$ ($k = 1, 2, \dots$) be disjoint sets and let $A = \bigcup_{k=1}^{\infty} A_k$. Let us put $B_p = \bigcup_{k=1}^p A_k$ for any natural p . Since $\mathfrak{S}$ is an algebra we have $B_p \in \mathfrak{S}$. Then for an arbitrary $E \subset X$ we must have

$$\mu^* E = \mu^*(E \cap B_p) + \mu^*(E \cap B_p')$$

Applying formula (6) (which is readily extended by induction to any finite number of disjoint μ^* -measurable sets) to the first summand in the right-hand member of the last relation and taking into account the monotonicity of μ^* we obtain

$$\begin{aligned} \mu^* E &= \sum_{k=1}^p \mu(E \cap A_k) + \mu^*(E \cap B_p') \geq \\ &\geq \sum_{k=1}^p \mu^*(E \cap A_k) + \mu^*(E \cap A') \end{aligned}$$

Next, passing to the limit as $p \rightarrow \infty$ and using the countable semiadditivity of the outer measure μ^* we see that

$$\mu^* E \geq \sum_{k=1}^{\infty} \mu^*(E \cap A_k) + \mu^*(E \cap A') \geq \mu^*(E \cap A) + \mu^*(E \cap A')$$

We have thus arrived at inequality (5), whence, as is known, it follows that A possesses property (4) for an arbitrary E , that is $A \in \mathfrak{S}$. Besides, putting $E = A$ we receive $\mu A \geq$

$$\geq \sum_{k=1}^{\infty} \mu A_k, \text{ and, since the opposite inequality is implied by}$$

the definition of outer measure, we obtain

$$\mu A = \sum_{h=1}^{\infty} \mu A_h$$

It now remains to check that $\mathfrak{S}$ is a σ -algebra. Let $A = \bigcup_{h=1}^{\infty} A_h$ where A_h are arbitrary sets belonging to $\mathfrak{S}$. Then A is representable by means of formula (4) in Chapter I as a union of disjoint sets and these sets belong to $\mathfrak{S}$. Consequently, by what was proved, $A \in \mathfrak{S}$.

The proof of the theorem has been completed.

We shall say that the measure μ constructed from a given outer measure μ^* in Theorem IV.3.2 is generated by the outer measure μ^* .

Let us enumerate some properties of μ^* -measurable sets.

1°. If $\mu^* A = 0$ then $A \in \mathfrak{S}$.

This property is quite obvious since, by virtue of the monotonicity of μ^* , inequality (5) holds for any $E \subset X$.

2°. If $A \in \mathfrak{S}$, $\mu A = 0$ and $E \subset A$ then $E \in \mathfrak{S}$.

Indeed, the monotonicity of μ^* implies that $\mu^* E = 0$, and it only remains to apply property 1°.

3°. If $A_1 \subset A \subset A_2$ and $A_1, A_2 \in \mathfrak{S}$ while $\mu A_1 = \mu A_2 < +\infty$ the set A also belongs to $\mathfrak{S}$ (besides, by the monotonicity of μ , we have $\mu A = \mu A_1 = \mu A_2$).

Indeed, $A \setminus A_1 \subset A_2 \setminus A_1$. But $A_2 \setminus A_1 \in \mathfrak{S}$ and $\mu(A_2 \setminus A_1) = 0$; therefore, by 2°, there must be $A \setminus A_1 \in \mathfrak{S}$ and hence $A = A_1 \cup (A \setminus A_1) \in \mathfrak{S}$.

4°. Criterion for μ^* -measurability. Let $E \subset X$. If, given any $\varepsilon > 0$, there exist $A, B \in \mathfrak{S}$ such that $A \subset E \subset B$ and $\mu(B \setminus A) < \varepsilon$ then $E \in \mathfrak{S}$.

Proof. For any natural n it is possible to choose $A_n, B_n \in \mathfrak{S}$ such that

$$A_n \subset E \subset B_n \quad \text{and} \quad \mu(B_n \setminus A_n) < \frac{1}{n}$$

Let us put $A = \bigcup_{n=1}^{\infty} A_n$ and $B = \bigcap_{n=1}^{\infty} B_n$. Then $A, B \in \mathfrak{S}$,

$A \subset E \subset B$ and $B \setminus A \subset B_n \setminus A_n$ for any n . Consequently, $\mu(B \setminus A) = 0$. According to 2°, it follows that $E \setminus A \in \mathfrak{S}$, and therefore we also have $E \in \mathfrak{S}$.

§ 4. Standard Extension of Measure

Combining the two procedures elaborated in the foregoing section we shall now construct an extension of a measure m in a set X defined on a semiring $\mathfrak{R}$, which leads to a new measure defined on a wider class of subsets of X . The extension process presented below was suggested by the German mathematician C. Carathéodory (1873-1950).

Theorem IV.4.1. *Let m be a measure in X defined on a semiring $\mathfrak{R}$ and let μ^* be the outer measure generated by the measure m and μ be the measure generated by the outer measure μ^* . Then μ is the extension of the measure m to the σ -algebra $\mathfrak{S}$ of μ^* -measurable sets, that is $\mathfrak{R} \subset \mathfrak{S}$ and $mA = \mu A$ for $A \in \mathfrak{R}$.*

In what follows the measure μ thus obtained will be referred to as the *standard (Carathéodory) extension of the measure m* ; the μ^* -measurable sets will also be called *m -measurable*.

Proof. We only need to verify the inclusion relation $\mathfrak{R} \subset \mathfrak{S}$ since the equality $mA = \mu A$ (for $A \in \mathfrak{R}$) will then follow from Theorem IV.3.1.

Let $A \in \mathfrak{R}$ and let E be an arbitrary subset of X . Let us check inequality (5); as was mentioned in the foregoing section, it is sufficient to consider the case when $\mu^*E < +\infty$.

Based on the definition of the outer measure μ^* generated by the given measure m , we see that, given an arbitrary $\varepsilon > 0$, there are $A_n \in \mathfrak{R}$ ($n = 1, 2, \dots$) such that

$$E \subset \bigcup_n A_n, \quad \sum_n mA_n < \mu^*E + \varepsilon$$

Further, we have the obvious inclusion relations

$$E \cap A \subset \bigcup_n (A_n \cap A) \quad \text{and} \quad E \cap A' \subset \bigcup_n (A_n \cap A')$$

The sets A_n satisfy the condition $A_n \cap A \in \mathfrak{R}$ and form a covering of the set $E \cap A$; consequently,

$$\mu^*(E \cap A) \leq \sum_n m(A_n \cap A) \quad (7)$$

The sets $A_n \cap A'$ are representable as the differences $A_n \setminus (A_n \cap A)$:

$$A_n \cap A' = A_n \setminus (A_n \cap A)$$

Therefore for every n there exist disjoint sets $C_{nk} \in \mathfrak{R}$ ($k = 1, 2, \dots$) such that $A_n \cap A' = \bigcup_k C_{nk}$. Besides,

$$\sum_k mC_{nk} = mA_n - m(A_n \cap A)$$

The system of sets $\{C_{nk}\}$ is a covering of the set $E \cap A'$, and consequently

$$\mu^*(E \cap A') \leq \sum_{n,k} mC_{nk} = \sum_n mA_n - \sum_n m(A_n \cap A) \quad (8)$$

Adding together (7) and (8) we obtain

$$\mu^*(E \cap A) + \mu^*(E \cap A') \leq \sum_n mA_n < \mu^*E + \varepsilon$$

By the arbitrariness of ε , this implies inequality (5), and the theorem has thus been proved.

Remark. If $X = \bigcup_{n=1}^{\infty} A_n$ where $A_n \in \mathfrak{R}$ and $mA_n < +\infty$ for any n , not only the measure m itself is σ -finite but also the standard extension μ of m possesses the same property. Obviously, the converse is also true: if μ is σ -finite, m is also σ -finite, and X admits of the above representation.

The simplest properties of m -measurable sets were enumerated at the end of the foregoing section. In particular, any subset E of an m -measurable set A with $\mu A = 0$ is also m -measurable (and $\mu E = 0$). In this connection we introduce the following general definition.

Definition. Let m be an arbitrary measure in X defined on a semiring $\mathfrak{R}$. We say that the measure m is *complete* if the relations $A \in \mathfrak{R}$, $mA = 0$ and $E \subset A$ imply $E \in \mathfrak{R}$.

Now we can state that *the standard extension of any measure m is always a complete measure.*

A general property of a complete measure is expressed by the following theorem.

Theorem IV.4.2. *Let the domain of definition $\mathfrak{R}$ of a complete measure m in X be a σ -algebra. If $E \subset X$ and, given any $\varepsilon > 0$, there is $A \in \mathfrak{R}$ such that $E \subset A$ and $mA < \varepsilon$, then $E \in \mathfrak{R}$ and $mE = 0$.*

Proof. For any natural n we choose a set $A_n \in \mathfrak{R}$ such that $E \subset A_n$ and $mA_n < \frac{1}{n}$. Let us put $A = \bigcap_{n=1}^{\infty} A_n$. Then

$A \in \mathfrak{R}$ (since $\mathfrak{R}$ is a σ -algebra) and $mA = 0$ by the monotonicity of measure. Now, since $E \subset A$ it only remains to resort to the completeness of the measure m .

Let us come back to the Carathéodory standard extension of an arbitrary measure m (from a semiring $\mathfrak{R}$). It should be noted that the standard extension may not be the minimal one in the sense that the σ -algebra $\mathfrak{S}$ of m -measurable sets may be wider than the σ -algebra generated by the semiring $\mathfrak{R}$. Such a situation will be demonstrated in the next chapter. The theorem below shows that a repeated application of the Carathéodory construction does not lead to a further extension.

Theorem IV.4.3. *Let m be a measure in X defined on a semiring $\mathfrak{R}$ and let μ be its standard extension. If μ^* and ν^* are, respectively, the outer measures generated by the measures m and μ then $\nu^*E = \mu^*E$ for any $E \subset X$.*

The consequence of this theorem is that the system of all μ -measurable sets coincides with that of all m -measurable sets, and therefore the standard extension of the measure μ coincides with μ itself.

Proof. Since μ is the extension of m (and, in particular, $\mathfrak{R} \subset \mathfrak{S}$) it is clear that $\nu^*E \leq \mu^*E$ for any $E \subset X$. Consequently, if $\nu^*E = \infty +$ then also $\mu^*E = +\infty$ and $\nu^*E = \mu^*E$.

Now, let $\nu^*E < +\infty$. Given an arbitrary $\varepsilon > 0$, there exists an at most countable system of sets $A_n \in \mathfrak{S}$ ($n = 1, 2, \dots$) such that $E \subset \bigcup_n A_n$ and

$$\sum_n \mu A_n < \nu^*E + \varepsilon$$

because the outer measure ν^* is generated by the measure μ . But we have $\mu A_n = \mu^*A_n$, and therefore, for every n , there is an at most countable system of sets $B_{nk} \in \mathfrak{R}$ such that $A_n \subset \bigcup_k B_{nk}$ and

$$\sum_k m B_{nk} < \mu A_n + \frac{\varepsilon}{2^n}$$

It follows that

$$\sum_{n,k} m B_{nk} < \nu^*E + 2\varepsilon$$

At the same time, $E \subset \bigcup_{n,k} B_{nk}$ and therefore

$$\mu^*E \leq \sum_{n,k} mB_{nk} < \nu^*E + 2\varepsilon$$

By the arbitrariness of ε , it follows that $\mu^*E \leq \nu^*E$, and the theorem has thus been proved.

§ 5. The Uniqueness of the Extension of Measure

If there is a measure m defined in a set X (on a semiring $\mathfrak{R}$) it is possible that, besides its standard extension μ to the σ -algebra $\mathfrak{S}$ of m -measurable sets, there exist some other extensions of this measure to some other σ -algebras. At the same time, under some conditions, it turns out that, within the σ -algebra $\mathfrak{S}$, the standard extension is unique. The exact statement of this result will be given later. For our further aims we need the following lemma.

Lemma IV.5.1. *Let $E \subset X$ and $\mu^*E < +\infty$, μ^* being the outer measure in X generated by the measure m . Then there exists a decreasing sequence of sets $H_n \in \mathfrak{S}$ ($n = 1, 2, \dots$) satisfying the following conditions:*

- (i) $E \subset H_n$ and $\mu H_n < +\infty$ for any n ;
- (ii) each H_n is representable in the form of an at most countable union of disjoint sets T_{nk} belonging to the semiring $\mathfrak{R}$ (i. e. $H_n = \bigcup_k T_{nk}$, $T_{nk} \in \mathfrak{R}$) and $T_{n_1k_1} \cap T_{n_2k_2} = \emptyset$ for $T_{n_1k_1} \neq T_{n_2k_2}$);
- (iii) If $H = \bigcap_{n=1}^{\infty} H_n$ then $\mu H = \mu^*E$ where μ is the standard extension of the measure m .

From condition (ii) it follows immediately that the sets H_n and, together with them, the set H are contained in the σ -algebra generated by the semiring $\mathfrak{R}$.

Proof. The definition of μ^* shows that for each n there is an at most countable family of sets $A_{nk} \in \mathfrak{R}$ covering E such that

$$\sum_k mA_{nk} < \mu^*E + \frac{1}{n}$$

Let us put $D_n = \bigcup_k A_{nk}$. By the properties of a semiring (see proposition (b) in Sec. I.7), D_n can be represented in

the form of a union of disjoint sets $C_{nh} \in \mathfrak{R}$: $D_n = \bigcup_h C_{nh}$. The sets D_n are contained in $\mathfrak{S}$ and

$$\sum_h mC_{nh} = \mu D_n \leq \sum_h mA_{nh} < \mu^*E + \frac{1}{n}$$

(the middle inequality follows from the countable semiadditivity of the measure μ).

Now we shall show that all the sets D_n can be replaced by some new sets H_n forming a decreasing sequence and possessing the same properties as D_n . To this end we put $H_1 = D_1$. Furthermore, we have $D_1 = \bigcup_h C_{1h}$ and $D_2 = \bigcup_l C_{2l}$ where both unions involve disjoint sets belonging to $\mathfrak{R}$. Next we construct the sets $B_{kl} = C_{1h} \cap C_{2l}$. These sets B_{kl} belong to $\mathfrak{R}$; they are disjoint and constitute a covering of the set E .* Let us put $H_2 = \bigcup_{h,l} B_{kl}$. It is readily seen that $H_2 = H_1 \cap D_2$, and thus $E \subset H_2 \subset H_1$. We can similarly construct $H_3 = H_2 \cap D_3$, etc.

Since all the sets H_n belong to $\mathfrak{S}$ we also have $H = \bigcap_{n=1}^{\infty} H_n \in \mathfrak{S}$, and, by Theorem IV.1.2, $\mu H = \lim \mu H_n$. But

$$\mu^*E \leq \mu H_n \leq \mu D_n < \mu^*E + \frac{1}{n}$$

and hence $\mu H = \mu^*E$.

Theorem IV.5.1. *Let μ be the standard extension of the measure m from the semiring $\mathfrak{R}$ to the σ -algebra $\mathfrak{S}$, the measure μ being σ -finite; if ν is a measure which is an extension of the measure m to a σ -algebra $\mathfrak{T} \supset \mathfrak{R}$ then $\mu A = \nu A$ for all $A \in \mathfrak{S} \cap \mathfrak{T}$. If, in addition to the foregoing conditions, ν is a complete measure then $\mathfrak{S} \subset \mathfrak{T}$.*

Proof. (a) Let us begin with the case when $A \in \mathfrak{S} \cap \mathfrak{T}$ and $\mu A < +\infty$. On constructing the sets H_n and H satisfying the conditions of the foregoing lemma (for $E = A$) we obtain $H_n, H \in \mathfrak{S} \cap \mathfrak{T}$,

$$\mu H_n = \sum_h \mu T_{nh} = \sum_h mT_{nh} = \sum_h \nu T_{nh} = \nu H_n$$

* If $x \in E$ then $x \in C_{1h}$ and $x \in C_{2l}$ for some h and l ; therefore $x \in B_{hl}$.

and

$$\nu H = \lim \nu H_n = \lim \mu H_n = \mu H = \mu A$$

Let us put $N = H \setminus A$. Then $N \in \mathfrak{S} \cap \mathfrak{T}$ and $\mu N = 0$. Given an arbitrary $\varepsilon > 0$, there are sets $D_p \in \mathfrak{R}$ ($p = 1, 2, \dots$) such that $N \subset \bigcup_p D_p$ and $\sum_p m D_p < \varepsilon$. It follows that we also have

$$\nu N \leq \sum_p \nu D_p = \sum_p m D_p < \varepsilon$$

and therefore $\nu N = 0$ whence

$$\nu A = \nu H - \nu N = \nu H = \mu A$$

Now let $A \in \mathfrak{S} \cap \mathfrak{T}$ and $\mu A = +\infty$. The measure μ being σ -finite, we have $X = \bigcup_{n=1}^{\infty} X_n$ where $X_n \in \mathfrak{S}$ and $\mu X_n < +\infty$ for each n . Furthermore, for each X_n there exists an at most countable covering by sets of finite m -measure belonging to $\mathfrak{R}$. Hence, the entire set X also possesses a covering of this type. Therefore, by virtue of proposition (b) in Sec. I.7, there exists a covering of X consisting of disjoint sets belonging to $\mathfrak{R}$ for which the values of the measure m are finite:

$$X = \bigcup_{n=1}^{\infty} K_n, \quad K_n \in \mathfrak{R} \text{ and } m K_n < +\infty \quad (n = 1, 2, \dots)$$

Let us put

$$A_n = A \cap K_n \quad (n = 1, 2, \dots)$$

Then $A_n \in \mathfrak{S} \cap \mathfrak{T}$, the sets A_n are disjoint, $A = \bigcup_{n=1}^{\infty} A_n$ and $\mu A_n < +\infty$. By what was proved above, we have $\nu A_n = \mu A_n$ for all n , and, consequently, there must also be

$$\nu A = \sum_{n=1}^{\infty} \nu A_n = \sum_{n=1}^{\infty} \mu A_n = \mu A$$

It should be noted that if $\mathfrak{B}$ is the σ -algebra generated by the semiring $\mathfrak{R}$ we have $\mathfrak{B} \subset \mathfrak{S} \cap \mathfrak{T}$, and therefore $\mu A = \nu A$ for all $A \in \mathfrak{B}$.

(b) Now let us suppose that the measure ν is complete, and let $A \in \mathfrak{S}$ and $\mu A < +\infty$. By the above lemma, there

is a set $H \in \mathfrak{B}$ such that $A \subset H$ and $\mu H = \mu A$. As before, we put $N = H \setminus A$; then $\mu N = 0$. Applying the lemma repeatedly we find a set $M \in \mathfrak{B}$ such that $N \subset M$ and $\mu M = 0$. Then $M \in \mathfrak{L}$, and, according to the note made after the proof of (a), we must have $\nu M = 0$. By the completeness of the measure ν , the set N belongs to $\mathfrak{L}$, and therefore we also have $A = H \setminus N \in \mathfrak{L}$.

If $\mu A = +\infty$ we represent A in the form $A = \bigcup_{n=1}^{\infty} A_n$ where $A_n \in \mathfrak{G}$ and $\mu A_n < +\infty$ for each n and readily see that $A \in \mathfrak{L}$.

The condition of σ -finiteness of the measure μ is essential in this theorem. We shall demonstrate its role by the following elementary example. Let X consist of two points a and b and let $\mathfrak{R}$ be the ring formed of the single-element subsets (a) and the empty set $\emptyset$ for which we put $m(a) = 1$ and $m\emptyset = 0$. The measure μ defined for all the subsets of X (which are (a) , (b) , $\emptyset$ and X) and assuming the values

$$\mu(a) = 1, \quad \mu(b) = +\infty, \quad \mu\emptyset = 0 \quad \text{and} \quad \mu X = +\infty$$

is a standard extension of the measure m . This measure μ is not σ -finite. On the other hand, we can, for instance, take the measure ν determined by the equalities

$$\nu(a) = 1, \quad \nu(b) = 1, \quad \nu\emptyset = 0, \quad \nu X = 2$$

The measure ν is also an extension of the measure m to the same algebra of all the subsets of X and it does not coincide with μ .

In conclusion we mention one more characteristic property of the standard extension μ of a measure m from a semiring $\mathfrak{R}$ which only takes place in the case when X can be covered by a countable system of sets $A_n \in \mathfrak{R}$ with $m A_n < +\infty$. As follows from Theorem IV.5.1, in this case μ is a complete measure in X defined on a σ -algebra $\mathfrak{L} \supset \mathfrak{R}$, which is an extension of the measure m and for which $\mathfrak{L}$ is the minimal possible σ -algebra serving as its domain of definition (as we know, in this case $\mathfrak{L} = \mathfrak{G}$).

The Lebesgue Measure in a Euclidean Space

§ 1. n -dimensional Parallelepipeds

In this chapter we shall apply the general method elaborated in Chapter IV to construct a measure in the Euclidean space R_n . To begin with, we shall consider some simpler sets in the space R_n serving as a generalization of the notion of an interval on the real line.

Let there be given n pairs of real numbers ' a_i and b_i ' ($i = 1, 2, \dots, n$) such that $a_i < b_i$ for each i . Here we admit that some of these numbers may be improper, that is we do not exclude the cases when $a_i = -\infty$ or $b_j = +\infty$ (or both) for some i 's and j 's. The collection Δ_0 of all points $x = (\xi_1, \xi_2, \dots, \xi_n) \in R_n$ whose coordinates satisfy the inequalities

$$a_i < \xi_i < b_i \quad (i = 1, 2, \dots, n)$$

is called an *open n -dimensional ("rectangular") parallelepiped*. The collection Δ^* of all points $x \in R_n$ whose coordinates satisfy the inequalities

$$a_i \leq \xi_i \leq b_i \quad (i = 1, 2, \dots, n)$$

is called a *closed (n -dimensional) parallelepiped*.

The definition of the metric in the space R_n directly implies that every closed (open) parallelepiped is a closed (open) set in R_n . It is clear that Δ_0 is the set of all interior points of Δ^* and that $\Delta^* = \bar{\Delta}_0$, i.e. Δ^* is the closure of the open parallelepiped Δ_0 . By *boundary points* of an open set are meant all its limit points not belonging to it; we can say that Δ^* is obtained from Δ_0 by adding to the latter all its boundary points.

Every set Δ obtained from Δ_0 by adding to it *some part* of the set of its boundary points is also called an *n -dimensional parallelepiped*. Thus, by a parallelepiped Δ we mean

any set satisfying the condition $\Delta_0 \subset \Delta \subset \Delta^*$. It is obvious that the set of all interior points of the parallelepiped Δ also coincides with Δ_0 .*

If $-\infty < a_i < b_i < +\infty$ for all i we shall say that Δ is a parallelepiped *with finite edges* (or *finite dimensions*). If at least one of the quantities a_i and b_i is equal to ∞ we shall say that the parallelepiped Δ has an *infinite edge*.

An arbitrary parallelepiped Δ will be denoted as

$$\Delta = \langle a_1, b_1; a_2, b_2; \dots; a_n, b_n \rangle$$

(or, briefly, $\Delta = \langle a, b \rangle$). If Δ is an open parallelepiped we shall also write

$$\Delta = (a_1, b_1; a_2, b_2; \dots; a_n, b_n)$$

(or briefly, $\Delta = (a, b)$). In the case of a closed parallelepiped Δ we shall use the notation

$$\Delta = [a_1, b_1; a_2, b_2; \dots; a_n, b_n]$$

(or $\Delta = [a, b]$).**

In the one-dimensional case a parallelepiped Δ of an arbitrary type is an interval $\langle a, b \rangle$ which can be open, closed or half-open. In what follows we shall retain the notation $\langle a, b \rangle$ of an interval on the line when the type of the interval is not stipulated.

In what follows we shall distinguish between the cases when two (or more) given parallelepipeds are *nonintersecting* (*disjoint*), that is having no points in common, and when they are *nonoverlapping*: in the latter case we shall mean that they have no *interior points* in common but may have common *boundary points* (although the words "nonintersecting" and "nonoverlapping" are often used synonymously). For instance, the intervals $[1, 2]$ and $[2, 3]$ intersect but have no interior points in common, i.e. are nonoverlapping.

It is readily seen that if two parallelepipeds overlap each other their intersection is also a parallelepiped. It

* It should be noted that in the case $n = 2$ ($n = 3$) the definition of an n -dimensional parallelepiped leads to a rectangle (rectangular parallelepiped) with edges parallel to the coordinate axes.

** In contrast to the notation used in classical mathematical analysis we write $\Delta = [a, b]$ even in the case when some of the quantities a_i and b_i turn into ∞ .

is also clear that if an open parallelepiped intersects some other parallelepiped these two parallelepipeds are necessarily overlapping, that is there are interior points common

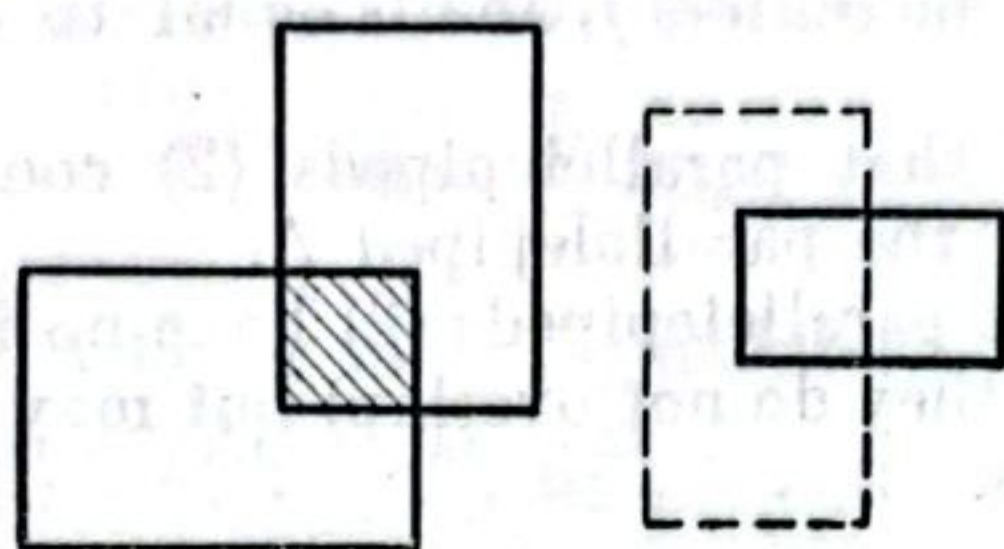


Fig. 9

to both parallelepipeds (see Fig. 9 where the two-dimensional case is shown)*.

Let us introduce the notion of a network partition of a parallelepiped $\Delta = \langle a, b \rangle$. Suppose that each of the inter-

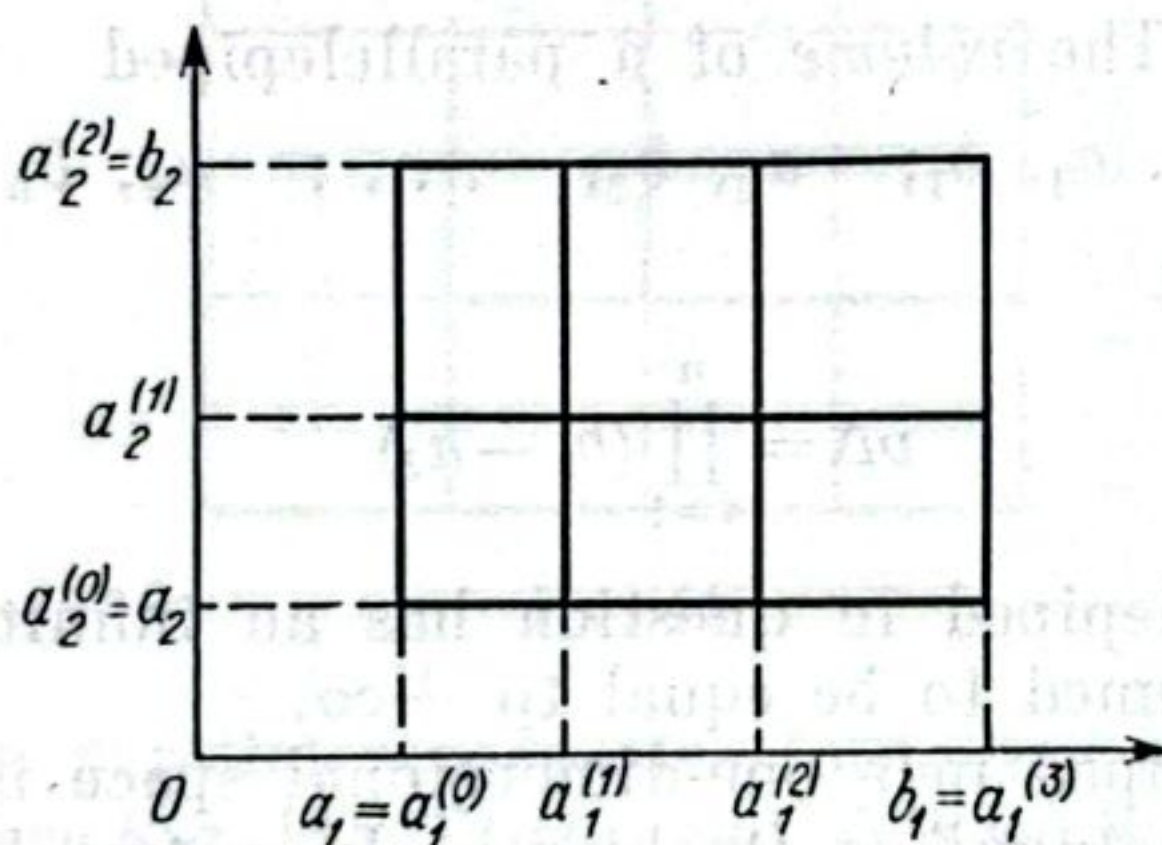


Fig. 10

vals $\langle a_i, b_i \rangle$ is broken up into a finite number of non-overlapping intervals with the aid of points of division

$$a_i = a_i^{(0)} < a_i^{(1)} < \dots < a_i^{(h_i)} = b_i \quad (1)$$

Let us form the parallelepipeds

$$\Delta_{j_1 j_2 \dots j_n} = \langle a_1^{(j_1)}, a_1^{(j_1+1)}; \dots; a_n^{(j_n)}, a_n^{(j_n+1)} \rangle \quad (2)$$

* If there is a common point of two parallelepipeds which is an interior point of one of them then in any neighbourhood of this point there are interior points of the other parallelepiped.

where each of the indices j_i can assume any value from 0 to $k_i - 1$ (each of the parallelepipeds can be of arbitrary type). The total number of these parallelepipeds depends on the ranges of the indices j_i and is equal to $k_1, k_2, \dots, k_n$ (see Fig. 10).

We shall say that parallelepipeds (2) constitute a *network partition* of the parallelepiped Δ .

It is clear that parallelepipeds (2) have no interior points in common (i.e. they do not overlap) but may have common boundary points.

§ 2. The Volume of a Parallelepiped

Generalizing the formula for the area of a rectangle in the plane and the formula for the volume of a rectangular parallelepiped in the three-dimensional space we introduce the following

Definition. The *volume* of a parallelepiped

$$\Delta = \langle a_1, b_1; a_2, b_2; \dots; a_n, b_n \rangle \quad (3)$$

is the product

$$v\Delta = \prod_{i=1}^n (b_i - a_i)$$

If the parallelepiped in question has an infinite edge its volume is assumed to be equal to $+\infty$.

A parallelepiped in a one-dimensional space is an interval and its "volume" is its length. In a two-dimensional space a parallelepiped is understood as a rectangle and its "volume" is its area.

The above formula for the volume immediately shows that if a system of parallelepipeds (2) forms a network partition of a parallelepiped (3) then

$$v\Delta = \sum_{j_1=0}^{k_1-1} \sum_{j_2=0}^{k_2-1} \dots \sum_{j_n=0}^{k_n-1} v\Delta_{j_1 j_2 \dots j_n}^* \quad (4)$$

Let us prove a number of lemmas.

* If $v\Delta = +\infty$ then at least one of the parallelepipeds (2) has a volume equal to $+\infty$.

Lemma V.2.1. *If a parallelepiped Δ is represented in the form of a finite union of pairwise nonoverlapping parallelepipeds Δ_k ($k = 1, 2, \dots, p$), that is $\Delta = \bigcup_{k=1}^p \Delta_k$, then*

$$v\Delta = \sum_{k=1}^p v\Delta_k$$

Proof.* Let Δ be of form (3) and let

$$\Delta_k = \langle c_1^{(k)}, d_1^{(k)}; c_2^{(k)}, d_2^{(k)}; \dots; c_n^{(k)}, d_n^{(k)} \rangle$$

$$(k = 1, 2, \dots, p)$$

Let us construct a network partition of the parallelepiped Δ into parallelepipeds of type (2) such that the collection

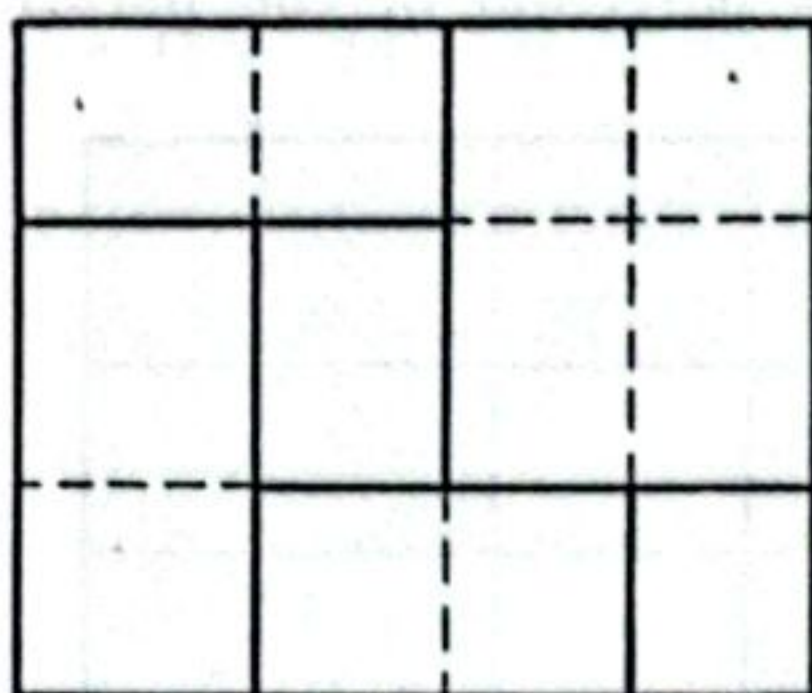


Fig. 11

of all these parallelepipeds also forms network partitions of all Δ_k 's (see Fig. 11). To this end, it is sufficient to arrange all the numbers $c_i^{(k)}$ and $d_i^{(k)}$ ($k = 1, 2, \dots, p$), taken for each i as a single increasing sequence. This results in sequences of type (1) which partition the intervals $\langle a_i, b_i \rangle$ into a finite number of subintervals. It is evident that parallelepipeds (2) constructed with the aid of these partitions possess the required properties.

Let T_k denote the system of all parallelepipeds of type (2) which form a network partition of the parallelepiped Δ_k ($k = 1, 2, \dots, p$). Each of the parallelepipeds (2) enters into one and only one of the collections T_k . From formu-

* For $n = 1$ and $n = 2$ this lemma and the following two lemmas are quite obvious.

la (4) it then readily follows that

$$v\Delta = \sum_{h=1}^p \sum_{\Delta_{j_1 j_2 \dots j_n} \in T_h} v\Delta_{j_1 j_2 \dots j_n} = \sum_{h=1}^p v\Delta_h$$

(the inner sum in the middle member of the equality extends over all the parallelepipeds entering into T_h).

Lemma V.2.2. *If the parallelepipeds Δ_k ($k = 1, 2, \dots, p$) do not overlap and are contained in the parallelepiped Δ , i.e. $\bigcup_{k=1}^p \Delta_k \subset \Delta$, then*

$$\sum_{k=1}^p v\Delta_k \leq v\Delta \quad (5)$$

The proof of the lemma is analogous to the above: we construct a network partition of the parallelepiped Δ into

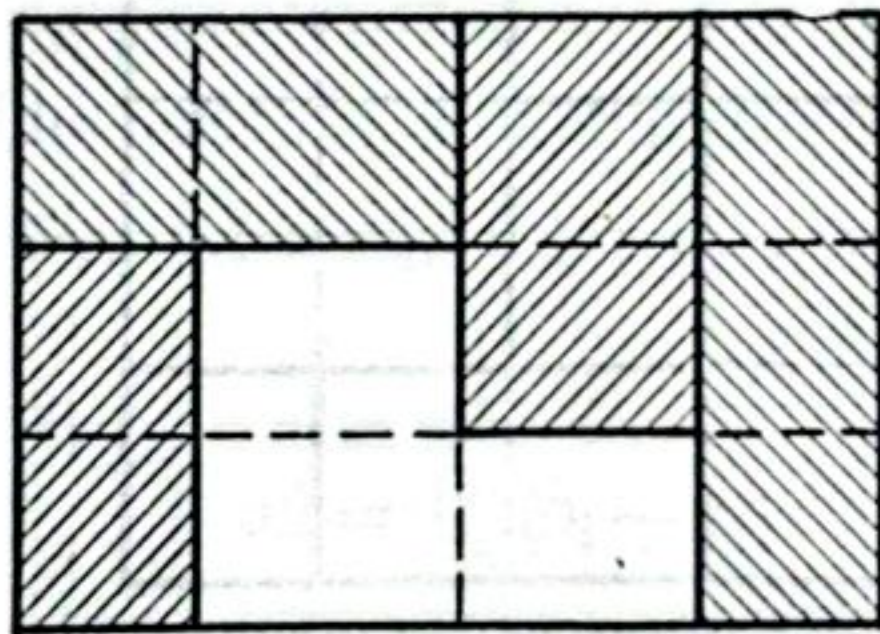


Fig. 12

parallelepipeds (2) in such a way that they simultaneously constitute network partitions of all Δ_k 's. Since in this construction none of the numbers $c_i^{(k)}$ and $d_i^{(k)}$ equals a_i or b_i we should add a_i and b_i to the collection of all $c_i^{(k)}$ and $d_i^{(k)}$ ($k = 1, 2, \dots, p$) to obtain the required sequences of points of division breaking up the intervals $\langle a_i, b_i \rangle$. Let the symbol T_k have the same sense as in the foregoing proof. Then each of the parallelepipeds (2) can belong to only one of the collections T_k but among (2) there may be parallelepipeds which do not belong to any T_k (Fig. 12). Therefore,

$$v\Delta \geq \sum_{h=1}^p \sum_{\Delta_{j_1 j_2 \dots j_n} \in T_h} v\Delta_{j_1 j_2 \dots j_n} = \sum_{h=1}^p v\Delta_h$$

Lemma V.2.3. *If parallelepipeds Δ and Δ_k ($k = 1, 2, \dots, p$) are such that $\Delta \subset \bigcup_{k=1}^p \Delta_k$ we have*

$$v\Delta \leq \sum_{k=1}^p v\Delta_k \quad (6)$$

(in this lemma the parallelepipeds Δ_k are allowed to overlap each other).

Proof. Since we can pass, without violating the conditions of the lemma, from the parallelepiped Δ to a new parallelepiped consisting of all the interior points of the former, it is allowable to assume, without loss of generality, that the given parallelepiped Δ is open. Besides, we can suppose that each of the Δ_k 's intersects Δ (and, consequently, overlaps Δ) because the addition of some extra summands to the right-hand side of inequality (6) only strengthens it.

Let us consider the parallelepipeds $\Delta'_k = \Delta_k \cap \Delta$ ($k = 1, 2, \dots, p$). We have $\Delta = \bigcup_{k=1}^p \Delta'_k$. Now, as in the two foregoing proofs, we construct a network partition of the parallelepiped Δ into parallelepipeds (2) which simultaneously serves as network partitions of all Δ'_k . If T_k denotes the collection of all parallelepipeds (2) which form the network partition of the parallelepiped Δ'_k then each of the parallelepipeds (2) enters into at least one collection T_k but may also belong to several T_k 's. Therefore formula (4) indicates that

$$v\Delta \leq \sum_{k=1}^p \sum_{\Delta_{j_1 j_2 \dots j_n} \in T_k} v\Delta_{j_1 j_2 \dots j_n} = \sum_{k=1}^p v\Delta'_k \leq \sum_{k=1}^p v\Delta_k$$

Remark. Let us dwell on parallelepipeds with *finite* edges. The formula for the volume implies that $v\Delta$ is a continuous function of the arguments a_i and b_i . If the numbers a'_i, a''_i, b'_i, b''_i are such that

$$a'_i < a_i < a''_i < b''_i < b_i < b'_i \quad (7)$$

and $\Delta' = (a', b')$ and $\Delta'' = [a'', b'']$ are, respectively, the corresponding open and closed parallelepipeds we have

the inclusion relation

$$\Delta'' \subset \Delta \subset \Delta' \quad (8)$$

Let us set an arbitrary $\varepsilon > 0$. By the continuity of the volume, it is possible to choose the numbers a'_i, a''_i and b'_i, b''_i so close, respectively, to a_i and b_i that (with the retention of inequalities (7)) we obtain

$$v\Delta' < v\Delta + \varepsilon \text{ and } v\Delta'' > v\Delta - \varepsilon \quad (9)$$

Thus, given an arbitrary $\varepsilon > 0$, for any parallelepiped Δ with finite edges there is an open parallelepiped $\Delta' \supset \Delta$ and a closed parallelepiped $\Delta'' \subset \Delta$ satisfying conditions (8) and (9).

Lemma V.2.4. *If a parallelepiped Δ is represented as a countable union of pairwise nonoverlapping parallelepipeds*

Δ_k , i.e. $\Delta = \bigcup_{k=1}^{\infty} \Delta_k$, then

$$v\Delta = \sum_{k=1}^{\infty} v\Delta_k \quad (10)$$

Proof. From Lemma V.2.2 it follows that inequality (5) holds for any finite p . Whence, on passing to the limit for $p \rightarrow \infty$, we obtain

$$\sum_{k=1}^{\infty} v\Delta_k \leq v\Delta \quad (11)$$

It is sufficient to establish the opposite inequality only for the case when $v\Delta_k < +\infty$ for all k since if $v\Delta_k = +\infty$ for at least one value of the index k inequality (11) immediately goes into equality (10).

Let us first suppose that $v\Delta < +\infty$. Given an arbitrary $\varepsilon > 0$, we can choose a closed parallelepiped $\Delta'' \subset \Delta$ and open parallelepipeds $\Delta'_k \supset \Delta_k$ ($k = 1, 2, \dots$) such that

$$v\Delta'' > v\Delta - \varepsilon, \quad v\Delta'_k < v\Delta_k + \frac{\varepsilon}{2^k} \quad (12)$$

The parallelepipeds Δ'_k form a covering of the parallelepiped Δ'' . By the Borel-Lebesgue Theorem II.6.1, this covering contains a finite subcovering:

$$\Delta'' \subset \Delta'_{k_1} \cup \dots \cup \Delta'_{k_p}$$

Therefore, by Lemma V.2.3, we have

$$v\Delta'' \leq v\Delta'_{k_1} + \dots + v\Delta'_{k_p} < \sum_{k=1}^{\infty} v\Delta'_k$$

It follows, by (12), that

$$v\Delta < v\Delta'' + \varepsilon < \sum_{k=1}^{\infty} v\Delta_k + 2\varepsilon$$

and therefore, since ε is arbitrary, there must be

$$v\Delta \leq \sum_{k=1}^{\infty} v\Delta_k \quad (13)$$

Inequalities (11) and (13) imply equality (10).

Let us proceed to the case $v\Delta = +\infty$. Given an arbitrary $\varepsilon > 0$, we choose open parallelepipeds Δ'_k satisfying the same conditions as before. Next we set an arbitrary natural number N . On replacing "the infinite edges" of the parallelepiped Δ by sufficiently large "finite edges" and giving "the other edges" sufficiently small decreases we can construct a closed parallelepiped $\Delta'' \subset \Delta$ such that $v\Delta'' > N$. As above, using the Borel-Lebesgue theorem we

see that $v\Delta'' < \sum_{k=1}^{\infty} v\Delta'_k$ whence $N < \sum_{k=1}^{\infty} v\Delta_k + \varepsilon$. Now, by the arbitrariness of ε and N , it becomes clear that $\sum_{k=1}^{\infty} v\Delta_k = +\infty$, that is we again arrive at equality (10).

It follows from Lemmas V.2.1 and V.2.4 that v is a *countably additive function* of Δ . But it is not a measure since the parallelepipeds do not form a ring.

§ 3. Semiring of Cells

To construct a measure in a Euclidean space it is most convenient to use parallelepipeds of a special type defined below.

Definition. A parallelepiped $\Delta = \langle a, b \rangle$ will be called an *n-dimensional cell* if it consists of all the points x whose coordinates satisfy the inequalities

$$a_i \leq \xi_i < b_i \quad (i = 1, 2, \dots, n)$$

To denote a cell we introduce the symbol

$$\Delta = [a_1, b_1; a_2, b_2; \dots; a_n, b_n)$$

(or, in the shortened notation, $\Delta = [a, b)$).

We can say, conditionally, that a cell is a parallelepiped which is "closed from the left" and "open from the right". In a one-dimensional space a cell is a half-open interval

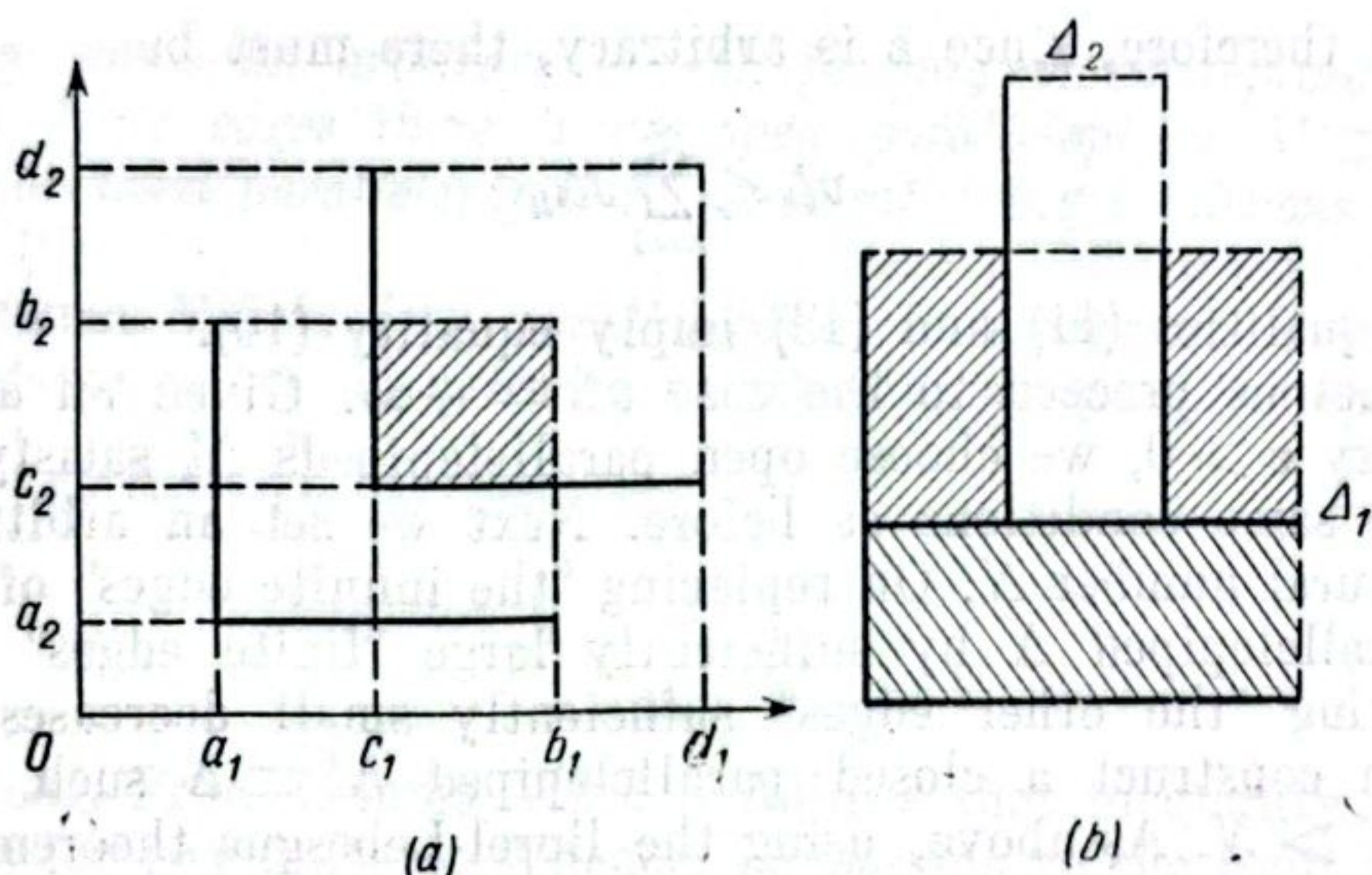


Fig. 13

including its left end point and not containing the right end point.

An empty set $\emptyset$ will also be regarded as a cell. The volume of an empty cell is put, by definition, equal to zero.

Let us establish some properties of cells.

(a) *The intersection of any two cells is also a cell.*

Proof. Indeed, if $\Delta_1 = [a, b)$ and $\Delta_2 = [c, d)^*$ then the intersection $\Delta_1 \cap \Delta_2$ is either empty or is representable as $\Delta_1 \cap \Delta_2 = [\alpha, \beta)$ where $\alpha_i = \max(a_i, c_i)$ and $\beta_i = \min(b_i, d_i)$ ($i = 1, 2, \dots, n$) (Fig. 13a).

(b) *The difference $\Delta_1 \setminus \Delta_2$ of two cells can be represented in the form of a finite union of disjoint cells (see Fig. 13b).*

* In accordance with the above convention, the symbol $[c, d)$ is the shortened notation of the cell

$$[c_1, d_1; c_2, d_2; \dots; c_n, d_n)$$

Proof. Since $\Delta_1 \setminus \Delta_2 = \Delta_1 \setminus (\Delta_1 \cap \Delta_2)$ it is allowable, without loss of generality, to assume, at the very beginning, that $\Delta_2 \subset \Delta_1$. Besides, we can suppose that $\Delta_2 \neq \emptyset$.

If $n = 1$ the assertion is obvious: the difference of two half-open intervals not including their right end points is a single half-open interval of the same type or a union of two such half-open intervals or an empty set. Now let us suppose that the assertion to be proved holds for $(n - 1)$ -dimensional cells. We shall prove that then it is valid for n -dimensional cells too.

We introduce the following notation: if Δ' is an $(n - 1)$ -dimensional cell and $a_n < b_n$ the symbol $\Delta' \times [a_n, b_n)$ designates the n -dimensional cell consisting of all the points $x = (\xi_1, \dots, \xi_{n-1}, \xi_n) \in \mathbf{R}_n$ for which

$$(\xi_1, \xi_2, \dots, \xi_{n-1}) \in \Delta' \text{ and } a_n \leq \xi_n < b_n$$

Let there be given two n -dimensional cells $\Delta_1 = [a, b)$ and $\Delta_2 = [c, d)$. If $a_n = c_n$ and $b_n = d_n$ we construct the $(n - 1)$ -dimensional cells

$$\Delta'_1 = [a_1, b_1; a_2, b_2, \dots, a_{n-1}, b_{n-1})$$

and

$$\Delta'_2 = [c_1, d_1; c_2, d_2; \dots, c_{n-1}, d_{n-1})$$

to represent Δ_1 and Δ_2 as

$$\Delta_1 = \Delta'_1 \times [a_n, b_n) \text{ and } \Delta_2 = \Delta'_2 \times [a_n, b_n)$$

It follows that

$$\Delta_1 \setminus \Delta_2 = (\Delta'_1 \setminus \Delta'_2) \times [a_n, b_n)$$

By the induction hypothesis, we have $\Delta'_1 \setminus \Delta'_2 = \bigcup_{h=1}^p \Delta''_h$ where Δ''_h are disjoint $(n - 1)$ -dimensional cells. Therefore,

$$\Delta_1 \setminus \Delta_2 = \bigcup_{h=1}^p \{\Delta''_h \times [a_n, b_n)\}$$

which is the required representation of the difference $\Delta_1 \setminus \Delta_2$.

Now suppose that $a_n < c_n < d_n < b_n$. In this case we take the cells

$$\Delta_3 = [c_1, d_1; \dots; c_{n-1}, d_{n-1}; a_n, c_n)$$

and

$$\Delta_4 = [c_1, d_1; \dots; c_{n-1}, d_{n-1}; d_n, b_n)$$

It is obvious that the cells Δ_2 , Δ_3 and Δ_4 are pairwise disjoint and that

$$\Delta_2 \cup \Delta_3 \cup \Delta_4 = [c_1, d_1; \dots; c_{n-1}, d_{n-1}; a_n, b_n) \subset \Delta_1$$

By what was proved above, the difference $\Delta_1 \setminus (\Delta_2 \cup \Delta_3 \cup \Delta_4)$ is representable in the form of a finite union of disjoint n -dimensional cells:

$$\Delta_1 \setminus (\Delta_2 \cup \Delta_3 \cup \Delta_4) = \bigcup_{k=1}^p \tilde{\Delta}_k$$

Therefore

$$\Delta_1 \setminus \Delta_2 = \Delta_3 \cup \Delta_4 \cup \bigcup_{k=1}^p \tilde{\Delta}_k$$

which is the desired representation.

In the case when $a_n < c_n < d_n = b_n$ (or $a_n = c_n < d_n < b_n$) the proof is carried out analogously with the only distinction that instead of the two cells Δ_3 and Δ_4 we use only one of them, namely Δ_3 (respectively, Δ_4).

The propositions proved show that *the collection of all cells in the space R_n is a semiring*. We shall denote it $\mathfrak{R}$.

Theorem V.3.1. *The function ν defined on the semiring $\mathfrak{R}$ (which for each cell assumes the value equal to the volume of that cell) is a σ -finite measure in R_n .*

Proof. By definition, $\nu\Delta \geq 0$ for any $\Delta \in \mathfrak{R}$. The σ -additivity of the function ν was proved in Sec. V.2. Finally, the space R_n can be represented as a countable union of cells of finite volume, for instance, in the form of the union of the cells

$$\Delta_p = [-p, p; -p, p; \dots; -p, p) \quad (p = 1, 2, \dots)^* \quad (14)$$

§ 4. Representation of an Open Set with the Aid of Cells

In Sec. II.5 we already established a theorem on the structure of open sets in the space R_n . This theorem is only a partial analogue of Theorem II.4.1^{*} describing the struc-

^{*} In other words, the cell Δ_p consists of all the points $x \in R_n$ for which $-p \leq \xi_i < p$ for any $i = 1, 2, \dots, n$.

ture of an open set on the real line since it asserts that an arbitrary open set in $\mathbf{R}_n$ can be represented in the form of a union of open balls (which are open sets of a simpler type) but these balls can intersect each other. Here we are going to prove that any open set in $\mathbf{R}_n$ can be partitioned (not uniquely) into (disjoint) sets of another simple type, namely, into disjoint cells. But the latter simpler sets are no longer open.

Theorem V.4.1. *Any open set $G \subset \mathbf{R}_n$ is representable in the form of an at most countable union of disjoint n -dimensional cells with finite edges*.*

Proof. For each natural number m we construct the partition of the space $\mathbf{R}_n$ into cells $[a, b)$ where every number a_i can take on an arbitrary value of the form $k/2^m$ where k is an integer ($k = 0, \pm 1, \pm 2, \dots$) and every b_i is equal to $a_i + \frac{1}{2^m}$. These cells will be referred to as the cells of the m th rank. It is clear that the cells of one rank are pairwise disjoint and that each cell of the $(m+1)$ th rank is contained entirely within one of the cells of the m th rank**.

Consider an arbitrary open set $G \subset \mathbf{R}_n$. Let us suppose that $G \neq \emptyset$; if otherwise the set G itself is a cell. Now, from the collection of the cells of the 1st rank we choose those which are entirely contained within G and denote this collection as Σ_1 (Σ_1 may also turn out to be empty)***. Next, from the collection of all the cells of the 2nd rank we choose those which are entirely contained in G but are not contained in any of the cells belonging to Σ_1 (and thus do not intersect them). We denote the collection of these cells of the 2nd rank as Σ_2 . In Fig. 14 the collection Σ_1 consists of 3 square cells with side $1/2$ and Σ_2 consists of 10 cells with side $1/4$. Let us continue this process indefinitely. In this construction the collection Σ_m includes all the cells of the m th rank entirely contained in G which

* It is apparent that if $G \neq \emptyset$ it cannot be represented as a finite union of cells.

** More precisely, the cells of the $(m+1)$ th rank appear as the result of the partition of each of the cells of the m th rank into 2^m parts.

*** Note that by Σ_1 we mean the *collection of the cells* but not the point set consisting of all the points of these cells.

are not contained in any of the cells entering into the collections $\sum_1, \sum_2, \dots, \sum_{m-1}$. Let H be the subset of R_n which is the union of all the cells entering into $\bigcup_{m=1}^{\infty} \sum_m$. Since the cells of any rank form a countable set, each of the collections $\sum_m$ is at most countable; therefore H is an

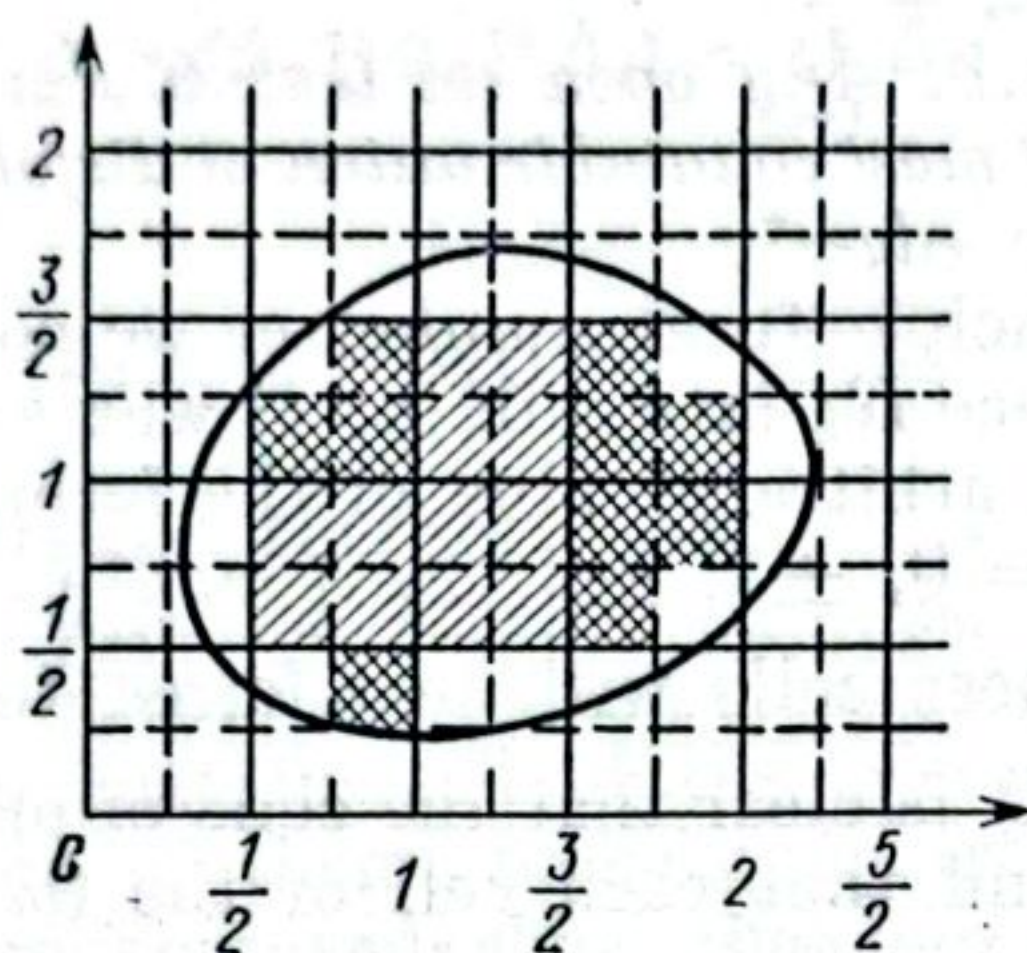


Fig. 14

at most countable union of cells. Besides, by construction, the cells included into the collection H are pairwise disjoint.

Let us prove that $G = H$. The inclusion relation $H \subset G$ is obvious from the construction; it only remains to establish the opposite inclusion relation. Let $x_0 \in G$. Then there is an open ball $S(x_0, \varepsilon) \subset G$. If m is chosen so that $\frac{\sqrt[n]{n}}{2^m} < \varepsilon$

the computations show that the cell of the m th rank which contains the point x_0 is itself entirely contained in $S(x_0, \varepsilon)$, and, consequently, in G as well. Let us consider those numbers m for which the cell of the m th rank containing x_0 is entirely contained in G (we have already shown that such values of m exist). Now, from these m 's we choose the least one; let it be $m = m_0$. Denoting by Δ_0 the cell of the m_0 th rank which contains x_0 we can assert that the cells of the m th rank with $m < m_0$ containing Δ_0 cannot be entirely contained in G and therefore they do not enter into $\sum_m$. Consequently, by construction, we have $\Delta_0 \in \sum_{m_0}$.

and therefore $\Delta_0 \subset H$ and $x_0 \in H$. Thus, the relation $G \subset H$ has been proved and, together with it, the equality $G = H$ has been established.

§ 5. Measurable Sets

Proceeding from the measure ν defined in Sec. V.3 for the semiring $\mathfrak{R}$, i.e. from the volume of a cell, and applying the process of (Carathéodory) extension of a measure described in Chapter IV we can construct the standard extension of the measure ν in the space R_n .

Definition. The standard extension μ of the volume ν is called the *Lebesgue measure* (in what follows we shall often simply call it the *measure* in R_n). The sets for which the measure μ is defined (that is the ν -measurable sets) are called *Lebesgue measurable* (or, simply, *measurable*).

In the space R_n we can of course also define various other measures but in this chapter the letter μ will denote the Lebesgue measure.

The measure μ is σ -finite since R_n is representable as a countable union of cells with finite volumes (see the proof of Theorem V.3.1). For all the sets belonging to the space R_n there is defined the outer measure μ^* , generated by the measure ν , which is spoken of as the *Lebesgue exterior* (also *outer*) *measure*. By construction, the measure μ is the restriction of the exterior measure μ^* to the σ -algebra $\mathfrak{S}$ of measurable sets.

Since the collection of all measurable sets forms a σ -algebra, the union and the intersection of a finite or countable number of measurable sets are measurable and the difference of two measurable sets is measurable; in particular, the complement of a measurable set with respect to the whole space R_n is also measurable.

The general properties of a measure imply that if $E \subset \bigcup_i E_i$ (in particular, if $E = \bigcup_i E_i$) where all the sets E_i and E are measurable and the union is finite or countable, then $\mu E \leq \sum_i \mu E_i$ (see Sec. IV.2, proposition (b)); this means that the measure is *countably semiadditive*. If $E = \bigcup_i E_i$ and

the sets E_i are disjoint we have $\mu E = \sum_i \mu E_i$ (the σ -additivity of the measure). Since the Lebesgue measure is obtained as the standard extension of the volume v we conclude (see Sec. IV.4) that it is *complete*, that is *any subset of a set of measure zero is measurable* (and is also of measure zero).

Since the Lebesgue measure is generated by the outer measure μ^* the general criterion for μ^* -measurability (proposition 4° in Sec. IV.3) applies to it and we can state that *if $E \subset \mathbb{R}_n$ and, given an arbitrary $\varepsilon > 0$, there are two measurable sets A and B such that $A \subset E \subset B$ and $\mu(B \setminus A) < \varepsilon$ the set E is also measurable*. In what follows we shall simply refer to the latter criterion as the *measurability criterion in $\mathbb{R}_n$* .

By its construction the σ -algebra of measurable sets includes all the cells and $\mu\Delta = v\Delta$ for any cell Δ .

For our further considerations the following remark is of use: *if it is known that the intersection of an arbitrary set $E \subset \mathbb{R}_n$ with the cells Δ_p specified by formula (14) is measurable for at least all sufficiently large values of p (then it is obviously measurable for all p as well) the set E is also measurable*.

Indeed, let us put $E_p = E \cap \Delta_p$ ($p = 1, 2, \dots$) and let E_p be measurable for all $p \geq p_0$. It is clear that $E = \bigcup_{p=p_0}^{\infty} E_p$, and therefore E is also measurable.

Theorem V.5.1. *Every open set and every closed set in $\mathbb{R}_n$ are measurable.*

Proof. The measurability of an open set follows from that of cells and from Theorem V.4.1. Since every closed set is the complement of an open set the former is also measurable.

Theorem V.5.2. *Any parallelepiped Δ is measurable and $\mu\Delta = v\Delta$.*

Proof. The measurability of open and closed parallelepipeds is implied by the foregoing theorem. Let us show that their measures coincide with their volumes.

Let us begin with an open or closed parallelepiped with finite edges. Given an arbitrary $\varepsilon > 0$, the argument similar to that in Sec. V.2 readily shows that there exist two cells Δ' and Δ'' for which conditions (8) and (9) hold. Then

$$v\Delta - \varepsilon < v\Delta'' = \mu\Delta'' \leq \mu\Delta \leq \mu\Delta' = v\Delta' < v\Delta + \varepsilon$$

whence $v\Delta - \varepsilon < \mu\Delta < v\Delta + \varepsilon$, and, by the arbitrariness of ε , the equality $\mu\Delta = v\Delta$ follows.

Now let $\Delta = \langle a, b \rangle$ be an arbitrary parallelepiped with finite edges. Introducing the parallelepipeds $\Delta_0 = (a, b)$ and $\Delta^* = [a, b]$ we can write $\Delta_0 \subset \Delta \subset \Delta^*$ and $\mu\Delta_0 = \mu\Delta^* = v\Delta$. According to proposition 3° in Sec. IV.3, it follows that Δ is measurable and $\mu\Delta = v\Delta$.

Lastly, let Δ be a parallelepiped having an infinite edge. Let us put $E_p = \Delta \cap \Delta_p$ where Δ_p are the cells determined by formula (14). For all sufficiently large p the parallelepiped Δ and the cell Δ_p overlap each other and therefore their intersection E_p is a parallelepiped (with finite edges). Consequently, the set E_p is measurable. Therefore, as was mentioned above, the parallelepiped Δ is measurable as well. It is clear that Δ also contains a cell Δ' with an infinite edge and, since $\mu\Delta \geq \mu\Delta' = +\infty$, we must have $\mu\Delta = +\infty$, that is $\mu\Delta = v\Delta$.

In particular, the theorem proved indicates that the measure of any interval in R_1 is equal to its length. Therefore Theorem II.4.1 shows that *the measure of a nonempty open set on the line is equal to the sum of the lengths of its components*.

Theorem V.5.3. *Every finite or countable set of points in R_n is measurable, its measure being equal to zero.*

Proof. A single-element set consisting of one point is closed and therefore measurable. Since it can be embedded into a parallelepiped of arbitrarily small volume its measure is equal to 0. The measurability of single-point sets implies that of any finite or countable point set. The measure of such a set is equal to zero since it is equal to the sum of the measures of the single-element sets it is formed of.

Theorem V.5.4. *All the Borel sets in R_n are measurable.*

Proof. The collection $\mathfrak{B}$ of all the Borel sets in R_n is the minimal σ -algebra containing all the closed subsets of R_n . The family $\mathfrak{S}$ of all the measurable sets in R_n is also a σ -algebra containing all the closed subsets of R_n . Consequently, $\mathfrak{B} \subset \mathfrak{S}$.

It should be noted that the σ -algebra $\mathfrak{S}$ is not exhausted by the Borel sets since there also exist measurable sets not belonging to $\mathfrak{B}$. This confirms that, as was mentioned in Sec. IV.4, a standard extension may not be the minimal

one. In the space R_n the minimal extension of the volume v from the semiring of cells is its extension to the σ -algebra $\mathfrak{B}$ of the Borel sets, that is the restriction of the Lebesgue measure to $\mathfrak{B}$.

Theorem V.5.5. *The exterior measure of any set $E \subset R_n$ is equal to the greatest lower bound of the measures of all the possible open sets G^* containing E :*

$$\mu^*E = \inf_{E \subset G} \mu G \quad (15)$$

Proof. From the monotonicity of the outer measure it follows that

$$\mu^*E \leq \mu G \quad \text{if } E \subset G \quad (16)$$

Therefore, if $\mu^*E = +\infty$ equality (15) is obvious. Let us assume that $\mu^*E < +\infty$. The outer measure μ^* being generated by the volume v , for any given $\varepsilon > 0$ there exists an at most countable system of cells Δ_k ($k = 1, 2, \dots$) such that

$$E \subset \bigcup_k \Delta_k, \quad \sum_k v\Delta_k < \mu^*E + \varepsilon$$

In particular, it follows that $v\Delta_k < +\infty$ for all k .

Let us choose for each k an open parallelepiped Δ'_k such that $\Delta_k \subset \Delta'_k$ and

$$\mu\Delta'_k < v\Delta_k + \frac{\varepsilon}{2^k}$$

Putting $G = \bigcup_k \Delta'_k$ we see that the set G is open, $E \subset G$ and

$$\mu G \leq \sum_k \mu\Delta'_k < \sum_k v\Delta_k + \varepsilon < \mu^*E + 2\varepsilon$$

Comparing the last relation with (16) and taking into account the arbitrariness of ε we conclude that equality (15) does in fact hold.

Corollary 1. *For any measurable set $E \subset R_n$ and for any $\varepsilon > 0$ there exist:*

(a) an open set $G \subset R_n$ such that $E \subset G$ and $\mu(G \setminus E) < \varepsilon$;

(b) a closed set $F \subset R_n$ such that $F \subset E$ and $\mu(E \setminus F) < \varepsilon$.

Proof. (a) If $\mu E < +\infty$ the result stated in (a) is an immediate consequence of Theorem V.5.5. In the general

* Open sets G of this kind always exist. For instance, such is $G = R_n$.

case we put

$$E_p = E \cap \Delta_p \quad (p = 1, 2, \dots)$$

where Δ_p are the cells specified by formula (14). By the above theorem, there are open sets $G_p \supset E_p$ such that $\mu(G_p \setminus E_p) < \varepsilon/2^p$. Let $G = \bigcup_{p=1}^{\infty} G_p$. Then the set G is open, $E \subset G$ and

$$G \setminus E \subset \bigcup_{p=1}^{\infty} (G_p \setminus E_p)$$

Consequently, $\mu(G \setminus E) < \varepsilon$.

(b) This part of the corollary follows from (a) if we pass to the complements. Indeed, let us put $E' = \mathbf{R}_n \setminus E$ and choose an open set $G \subset \mathbf{R}_n$ such that $E' \subset G$ and $\mu(G \setminus E') < \varepsilon$. The set $F = \mathbf{R}_n \setminus G$ is closed; besides, we have $F \subset E$ and $E \setminus F = G \setminus E'$. Consequently $\mu(E \setminus F) < \varepsilon$.

Corollary 2. *The measure of any measurable set $E \subset \mathbf{R}_n$ is equal to the least upper bound of the measures of all the possible closed sets F contained in E :*

$$\mu E = \sup_{F \subset E} \mu F \quad (17)$$

Proof. By analogy with (16), we have

$$\mu F \leq \mu E \quad \text{if } F \subset E$$

Furthermore, if $\mu E < +\infty$, the foregoing corollary implies that there exists a closed set $F \subset E$ whose measure is arbitrarily close to μE whence follows formula (17).

If $\mu E = +\infty$ and the set F satisfies condition (b) the previous corollary shows that we also have $\mu F = +\infty$, and formula (17) becomes quite apparent.

Theorem V.5.6. *For any measurable set $E \subset \mathbf{R}_n$ there is an F_σ -set H and a G_δ -set K (see Sec. II.8) such that*

$$H \subset E \subset K, \quad \mu H = \mu E = \mu K \quad \text{and} \quad \mu(K \setminus H) = 0$$

Proof. By Corollary 1 (a) of Theorem V.5.5, for every natural m there exists an open set $G_m \supset E$ such that $\mu(G_m \setminus E) < 1/m$. Let us put $K = \bigcap_{m=1}^{\infty} G_m$. Then K is a G_δ -set, $E \subset K$ and $\mu(K \setminus E) \leq \mu(G_m \setminus E) < 1/m$ for any m ; consequently $\mu(K \setminus E) = 0$. Similarly, assertion (b) of the same corollary implies the existence of an F_σ -set H

for which $H \subset E$ and $\mu(E \setminus H) = 0$. It follows that

$$\mu(K \setminus H) = \mu(K \setminus E) + \mu(E \setminus H) = 0$$

$$\mu E = \mu H + \mu(E \setminus H) = \mu H$$

and

$$\mu K = \mu E + \mu(K \setminus E) = \mu E *$$

We have incidentally proved that *every measurable set is representable as the union of a Borel F_σ -set and a measurable set of measure zero.*

In conclusion we mention without proof some other important facts. A mapping of the space $\mathbf{R}_n$ onto itself is said to be *isometric (length preserving)* if the distance between any two points is equal to the distance between the images of these points under the mapping. Two sets E_1 and E_2 belonging to $\mathbf{R}_n$ are called *congruent* if one of them is the image of the other under an isometric mapping. It can be proved that if two sets E_1 and E_2 are congruent and one of them is measurable the other is also measurable and $\mu E_1 = \mu E_2$. In other words, *the Lebesgue measure is invariant with respect to the isometric mappings.*

It is known that in each space $\mathbf{R}_n$ there exist sets which are not Lebesgue measurable. Moreover, it was proved that in any space $\mathbf{R}_n$ it is impossible to construct a σ -finite measure m such that (α) m is defined for all the subsets of $\mathbf{R}_n$; (β) m is invariant with respect to the length-preserving mappings, and (γ) the measure m of any parallelepiped coincides with its volume. The condition of σ -additivity which we included in the definition of a measure plays an essential role in the above proposition. If the concept of a measure is slightly generalized by admitting that the measure should only be finitely additive then, as was proved by S. Banach, in the spaces $\mathbf{R}_1$ and $\mathbf{R}_2$ there exist measures possessing properties (α)-(γ). But in case $n \geq 3$ even a finitely additive measure in $\mathbf{R}_n$ satisfying conditions (α)-(γ) does not exist. The latter result was established by the German mathematician F. Hausdorff (1868-1942).

* Thus, the inclusion relation $H \subset E \subset K$ and equality $\mu(K \setminus H) = 0$ automatically imply the equality $\mu H = \mu E = \mu K$ (for measurable sets). At the same time, it should be noted that the relation $\mu(K \setminus H) = 0$ only follows from the equality $\mu H = \mu K$, for $H \subset K$, when $\mu H < +\infty$.

Measurable Functions

§ 1. The Definition of a Measurable Function

In this chapter we shall study a class of functions playing a fundamental role in the definition of the integral given in the following chapter. The general facts of the theory will be presented for functions defined on an abstract measure space. Most of the results will be established for an arbitrary measure and only some of them will be proved under the additional requirement that the measure should be complete. Since the Lebesgue measure in the Euclidean space R_n is complete all the results proved in this chapter for functions on an abstract set will apply to R_n as well. At the end of the chapter we consider the special case of measurable functions in a Euclidean space.

Let X be an arbitrary set, $\mathcal{S}$ be a σ -algebra of its subsets, and μ be a measure defined on $\mathcal{S}$. The sets belonging to $\mathcal{S}$ will be called *measurable*. In particular, X may coincide with R_n , as $\mathcal{S}$ we can take the σ -algebra of all Lebesgue measurable sets and as μ the Lebesgue measure in R_n .

We shall begin with real functions with finite values whose domain of definition can be an arbitrary subset of X . As a rule, in what follows this subset will be supposed measurable.

Definition 1. Let a function f be defined on a set $E \subseteq X$. By the *Lebesgue sets* of this function are meant all the sets of the following four types*:

- (i) $E[f(x) > a]$,
- (ii) $E[f(x) \geq a]$,
- (iii) $E[f(x) < a]$,
- (iv) $E[f(x) \leq a]$,

* See the notation introduced in Sec. II.7.

where a can be any real number (see Fig. 15 demonstrating the graph of a function defined in the one-dimensional space $\mathbf{R}_1$ and its Lebesgue sets of types (i) and (ii)).

Definition 2. A function f defined on a set $E \subset X$ is said to be *measurable* (on this set) or, more precisely, $\mathfrak{S}$ -*measurable*, if all the Lebesgue sets of the function f belonging to

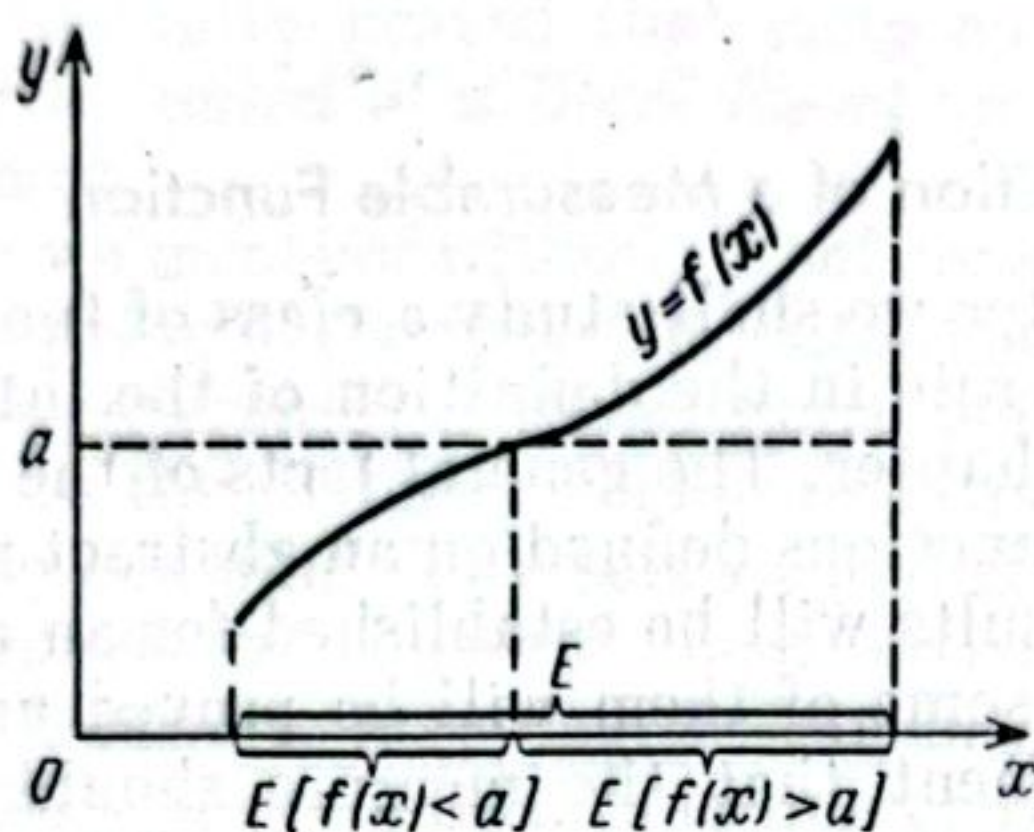


Fig. 15

the four types enumerated in Definition 1 are measurable for any a .

If $X = \mathbf{R}_n$ and $\mathfrak{S}$ consists of the Lebesgue measurable sets the measurable function is called *Lebesgue measurable*.

Note that if a function f is measurable on a set E the set E itself is measurable. This follows from the obvious equality

$$E = \bigcup_{n=1}^{\infty} E[f(x) > -n] \quad (1)$$

since the sets under the sign of union are Lebesgue's sets of the function f .

If f is measurable on E the set $E[f(x) = a]$ is also measurable for any a since it is representable as the intersection of two measurable sets:

$$E[f(x) = a] = E[f(x) \geq a] \cap E[f(x) \leq a]$$

Theorem VI.1.1. If a function f defined on a set E is such that its Lebesgue sets of one of the four types are measurable for all a this function is measurable.

Proof. Let, for instance, the Lebesgue sets $E[f(x) > a]$ of the first type be measurable for all a . We must show that

the Lebesgue sets of f belonging to the other three types are measurable as well.

For any a we have

$$E[f(x) \geq a] = \bigcap_{n=1}^{\infty} E\left[f(x) > a - \frac{1}{n}\right] \quad (2)$$

Indeed, the left-hand member of the equality is obviously contained in the right-hand member. Furthermore, if x belongs to the intersection entering into the right-hand side of formula (2) we have $f(x) > a - \frac{1}{n}$ for any n , which implies that $f(x) \geq a$. Hence, we have shown that the right-hand member is contained in the left-hand one whence follows equality (2). This equality and the measurability of all the Lebesgue sets of the function f belonging to the first type imply the measurability of the set $E[f(x) \geq a]$ and thus the Lebesgue sets of the second type are measurable.

As was shown above, formula (1) implies that if the Lebesgue sets of the first type are measurable the set E is itself measurable. Therefore the obvious relations

$$E[f(x) < a] = E \setminus E[f(x) \geq a]$$

and

$$E[f(x) \leq a] = E \setminus E[f(x) > a]$$

show that the measurability of the Lebesgue sets of the third and fourth types follows from the measurability of the Lebesgue sets of the first two types.* The measurability of f is similarly proved under the assumption that all its Lebesgue sets of any type other than the first are measurable. It should also be noted that the measurability of the set E can be proved on the basis of the measurability of the Lebesgue sets of any other type by using formulas similar to formula (1).*

Let us prove a number of simple propositions.

1°. If $f(x) = c$ (where c is a constant) on a measurable set E the function f is measurable.

* If, for instance, it is known that all the Lebesgue sets of the fourth type are measurable one should use the equality

$$E = \bigcup_{n=1}^{\infty} E[f(x) \leq n]$$

Proof. Since $f(x)$ is equal to the constant c we have

$$E[f(x) > a] = \begin{cases} E & \text{for } a < c \\ \emptyset & \text{for } a \geq c \end{cases}$$

whence follows that all the Lebesgue sets of the first type of the function f are measurable.

Let a function f be defined on a set $E \subset X$ and let $E' \subset E$. If all the Lebesgue sets of the function f formed for the set E' are measurable we say that the function f is measurable on E' .

2°. If f is measurable on a set E it is also measurable on every measurable subset $E' \subset E$.

Proof. For the Lebesgue sets of the first type of the function f on the set E' we have

$$E'[f(x) > a] = E' \cap E[f(x) > a]$$

whence follows their measurability.

3°. Let a function f be defined on a set E representable as a finite or countable union of sets E_i ($E = \bigcup_i E_i$). If f is measurable on each E_i it is also measurable on E .

Proof. For any a we have

$$E[f(x) > a] = \bigcup_i E_i[f(x) > a]$$

and, since each of the sets on the right-hand side is measurable, their union is measurable as well.

A special case of the last proposition reads as follows.

4°. Let $E = \bigcup_i E_i$ where all E_i 's ($i = 1, 2, \dots$) are measurable. If a function f defined on E assumes a constant value on each E_i then f is measurable on E .*

Indeed, this assertion immediately follows from 1° and 3°.

Definition. Let $E \subset X$. The function χ_E determined by the formulas

$$\chi_E(x) = \begin{cases} 1 & \text{for } x \in E \\ 0 & \text{for } x \in X \setminus E \end{cases}$$

* A measurable function of this kind taking no more than countably many distinct values is sometimes called a *simple function*.—Tr.

(and thus defined throughout X) is called the *characteristic* (or *defining*) *function* of the set E .

Proposition 4° obviously implies

5°. If E is a measurable set its characteristic function χ_E is measurable (on X).

It is easily proved that the converse is also true.

6°. If μ is a complete measure any function f defined on a set E with $\mu E = 0$ is measurable.

Proof. The measure μ being complete, any subset of the set E is measurable and therefore all the Lebesgue sets of the function f are measurable.

The theorem below pertains to functions in a Euclidean space.

Theorem VI.1.2. If a function f is continuous on a closed set $F \subset \mathbb{R}_n$ it is Lebesgue measurable on this set.

Proof. By Theorem II.7.1, if f is continuous on F all its Lebesgue sets of the second type are closed and, consequently, measurable (see Theorem V.5.1). Now it only remains to resort to Theorem VI.1.1.

Corollary. If f is a continuous function on a measurable set $E \subset \mathbb{R}_n$ it is measurable on E .

Proof. As is known from Sec. V.5, the set E can be represented as a union $E = E_1 \cup E_2$ where E_1 is an F_σ -set and $\mu E_2 = 0$. The set E_1 is representable in the form $E_1 = \bigcup_{k=1}^{\infty} F_k$ where the sets F_k are closed. By the theorem proved above, the function f is measurable on each set F_k and therefore its measurability on E_1 is implied by proposition 3°. On the other hand, according to proposition 6°, the function f is measurable on E_2 and therefore it is measurable on the entire set E .

In the above derivation of the simplest properties of measurable functions all the propositions, except proposition 6°, are only based on the fact that the measurable sets in question form a σ -algebra, while the values of the measure μ do not play any role. Therefore in the definition of a measurable function we may only suppose that there is an arbitrary σ -algebra $\mathcal{S}$ of subsets of the set X (which are called *measurable*); then it is natural to speak of $\mathcal{S}$ -measurable functions. For instance, if the space $\mathbb{R}_n$ is again taken as the set X while the σ -algebra $\mathfrak{B}$ of all the Borel sets in $\mathbb{R}_n$

is taken as $\mathfrak{S}$, definition 2 leads to the notion of $\mathfrak{B}$ -measurable functions. The $\mathfrak{B}$ -measurable functions in R_n are also termed the *Baire functions* after L. R. Baire (1874-1932), a French mathematician who approached this class of functions from a different point of view. Theorem V.5.4 shows that all the *Baire functions* are *Lebesgue measurable*.

The proof of Theorem VI.1.1 readily indicates that every function continuous on a closed subset of R_n is $\mathfrak{B}$ -measurable on this subset.

§ 2. Arithmetical Operations on Measurable Functions

In this section we shall show that the class of measurable functions is closed with respect to the arithmetical operations, which means that the operations of addition, subtraction, multiplication and division applied to measurable functions again result in measurable functions.

Lemma VI.2.1. *If a function f is measurable on a set $E \subset X$ the functions kf (where k is an arbitrary constant), $|f|$ and f^2 are measurable on E . If $f(x) \neq 0$ on E the function $1/f$ is also measurable on E .*

Proof. (a) Let us consider the function kf . If $k = 0$ we have $kf(x) \equiv 0$, and therefore the product kf is measurable as a constant (the measurability of the set E follows from the measurability of the function f). If $k \neq 0$ the measurability of the Lebesgue sets (of the first type) of the product kf is implied by the measurability of the Lebesgue sets of the function f and by virtue of the following obvious equalities:

$$E[kf(x) > a] = E\left[f(x) > \frac{a}{k}\right] \quad \text{for } k > 0$$

and

$$E[kf(x) > a] = E\left[f(x) < \frac{a}{k}\right] \quad \text{for } k < 0$$

(b) For the function $|f|$ we have

$$E[|f(x)| > a] = E \quad \text{if } a < 0$$

$$E[|f(x)| > a] = E[f(x) > a] \cup E[f(x) < -a] \quad \text{if } a \geq 0$$

whence it follows that all the Lebesgue sets of the first type of the function $|f|$ are measurable.

(c) For the Lebesgue sets of the function f^2 we can write

$$E[f^2(x) > a] = \begin{cases} E & \text{if } a < 0 \\ E[|f(x)| > \sqrt{a}] & \text{if } a \geq 0 \end{cases}$$

Since $|f|$ is measurable all these sets are measurable.

(d) Let $f(x) \neq 0$ at all the points of the set E . It is readily seen that

$$E\left[\frac{1}{f(x)} > a\right] = \begin{cases} E[f(x) > 0] \cap E\left[f(x) < \frac{1}{a}\right] & \text{if } a > 0 \\ E[f(x) > 0] \cup E\left[f(x) < \frac{1}{a}\right] & \text{if } a < 0 \\ E[f(x) > 0] & \text{if } a = 0 \end{cases}$$

In all the three cases the Lebesgue sets of the function $\frac{1}{f}$ are measurable.

Theorem VI.2.1. *If f and g are two measurable functions on a set E the functions $f \pm g$ and fg are measurable on E . If $g(x) \neq 0$ on E the function f/g is also measurable on E .*

Proof. (a) Let us arrange all rational numbers as a sequence $r_1, r_2, \dots, r_k, \dots$. This is possible because of the countability of the set of all rational numbers. We shall show that, for any real number a , there holds the equality

$$E[f(x) + g(x) > a] = \bigcup_{k=1}^{\infty} \{E[f(x) > a - r_k] \cap E[g(x) > r_k]\} \quad (3)$$

Let a point x belong to the left-hand member of equality (3). This means that $f(x) + g(x) > a$, i.e. $g(x) > a - f(x)$. By the density property of the set of all rational numbers, there is r_k such that $g(x) > r_k > a - f(x)$, that is $g(x) > r_k$ and $f(x) > a - r_k$. Thus, the point x belongs to one of the sets on the right-hand side of equality (3) and, consequently, it is contained in the right-hand member of that equality.

Conversely, if a point x belongs to the right-hand member of equality (3) we have $f(x) > a - r_k$ and $g(x) > r_k$ for some k . Consequently, $f(x) + g(x) > a$, that is x is contained in the left-hand member of equality (3) which has thus been proved.

The measurability of the function $f + g$ now readily follows from formula (3) and the measurability of the Lebesgue sets of the functions f and g .

(b) The difference $f(x) - g(x)$ can be written as

$$f(x) - g(x) = f(x) + [-g(x)]$$

By the lemma, the function $-g = (-1)g$ is measurable, which implies the measurability of $f - g$ as the sum of two measurable functions.

(c) The measurability of the product fg is proved with the aid of the equality

$$f(x)g(x) = \frac{1}{2} \{ [f(x) + g(x)]^2 - f^2(x) - g^2(x) \}$$

if we take into account what has been proved in (a) and (b) and use the lemma.

(d) If $g(x) \neq 0$ on E we have

$$\frac{f(x)}{g(x)} = f(x) \cdot \frac{1}{g(x)}$$

the function $1/g$ being measurable by the lemma and the measurability of the quotient f/g following from the measurability of the product of measurable functions.

Corollary. *If f and g are measurable on a set E both sets*

$$A = E[f(x) = g(x)] \quad \text{and} \quad B = E[f(x) \neq g(x)]$$

are measurable.

Indeed,

$$A = E[f(x) - g(x) = 0], \quad B = E \setminus A$$

§ 3. Limit in the Class of Measurable Functions

In this section we shall show that the class of measurable functions is closed with respect to the operation of passing to the limit, that is the limit function of a convergent sequence of measurable functions is measurable as well.

Lemma VI.3.1. *If f_k are measurable functions on a set $E \subset X$ whose number is finite or countable and the function $f(x) = \sup_k f_k(x)$ ($g(x) = \inf_k f_k(x)$) assumes finite values at*

all the points $x \in E$, the function f (respectively, g) is also measurable on E .

Proof. The measurability of the set $A = E [f(x) > a]$ for any a follows from the obvious equality $A = \bigcup_k A_k$ where $A_k = E [f_k(x) > a]$. Similarly, the measurability of the set $B = E [g(x) < a]$ is implied by the representation

$$B = \bigcup_k B_k \quad \text{where} \quad B_k = E [f_k(x) < a]$$

Theorem VI.3.1. *If functions f_k ($k = 1, 2, \dots$) and f are defined on a set $E \subset X$, the functions f_k being measurable on E for all k , and $f(x) = \lim f_k(x)$ for each $x \in E$, the function f is also measurable on the set E .*

Proof. Let us define on the set E the functions

$$\varphi_k(x) = \sup \{f_k(x), f_{k+1}(x), \dots\} \quad (k = 1, 2, \dots)$$

which are measurable by virtue of the foregoing lemma (the convergence of the sequence $\{f_k(x)\}$ implies that it is bounded at each point). The measurability of f now follows from the same lemma and the well-known relation

$$f(x) = \inf_k \varphi_k(x)^* \quad (4)$$

Corollary. *If u_k ($k = 1, 2, \dots$) are measurable functions on a set E , s is a function defined on the set E and the limiting relation*

$$s(x) = \sum_{k=1}^{\infty} u_k(x)$$

holds for all $x \in E$, the function s is also measurable on E .

* In other words, we use the fact that the limit of a sequence (provided it exists) coincides with its limit superior: $f(x) = \overline{\lim} f_k(x)$.

Formula (4) is proved as follows. Given an arbitrary $\varepsilon > 0$ and a fixed $x \in E$, there is k_0 such that

$$f(x) - \varepsilon < f_k(x) < f(x) + \varepsilon \quad \text{for } k \geq k_0$$

whence

$$f(x) - \varepsilon < \varphi_k(x) \leq f(x) + \varepsilon \quad \text{for } k \geq k_0$$

On the other hand, the definition of the functions φ_k implies that they form a decreasing sequence, and therefore $\inf \varphi_k(x) = \lim \varphi_k(x)$. On passing to the limit in the foregoing inequality we obtain

$$f(x) - \varepsilon \leq \inf \varphi_k(x) \leq f(x) + \varepsilon$$

which, by the arbitrariness of ε , implies equality (4).

The corollary is an obvious consequence of Theorems VI.2.1, VI.3.1 and of the definition of the sum of a series.

Let us consider functions defined on a closed set $F \subset \mathbb{R}_n$. As is already known, all continuous functions are measurable on F . From classical mathematical analysis it is also known that if a sequence of continuous functions is uniformly convergent the limit function is also continuous.* But if a sequence of continuous functions converges at each point but not uniformly the limit function is no longer necessarily continuous. According to the classification of functions introduced by Baire, a discontinuous function serving as the limit of a sequence of continuous functions is said to be of *the first class* (*Baire class 1*). By the above theorem, the functions of the first class are measurable (and even $\mathfrak{B}$ -measurable).

This classification of discontinuous functions can be continued as follows. A function not belonging to the Baire class 1 and representable as a pointwise limit of a sequence of functions of class 1 is of *Baire class 2*. A simple example of a function of Baire class 2 in the space $\mathbb{R}_1$ is the well-known *Dirichlet function*:

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational} \end{cases}$$

This function can be obtained from continuous functions by means of a two-fold passage to the limit:

$$f(x) = \lim_{m \rightarrow \infty} \lim_{h \rightarrow \infty} \cos^{2h}(m! \pi x)$$

Similarly, proceeding from functions of class 2 we can define functions of Baire class 3, etc. All the functions obtained with the aid of such subsequent passages to the limit turn out to be $\mathfrak{B}$ -measurable (i.e. they are Baire functions). But the family of all functions obtained from continuous functions by means of an arbitrary finite number of passages to the limit does not exhaust the class of all Baire functions. The latter class, in its turn, is essentially narrower than the class of all Lebesgue measurable functions.

* This assertion holds for functions defined on an arbitrary set in $\mathbb{R}_n$.

It is sometimes expedient to assume that the functions in question can take the values $+\infty$ and $-\infty$. The repetition of the definition given in Sec. VI.1 leads to the definition of measurability for such functions as well.*

It can easily be verified that the theorems and propositions established in this chapter continue to hold for functions assuming infinite values as well, with the only stipulation that when we speak of operations on functions it is required that the operations should make sense; for instance, for the addition of two functions we must require that there should be no points at which these functions assume infinite values of opposite signs. The only essential exception to this rule is Theorem VI.1.1 whose statement should additionally involve the condition that the set E (on which the function in question is defined) is measurable. The matter is that although, as before, the measurability of a function assuming infinite values implies the measurability of the set E in this theorem, the measurability of the set E (in contrast to the case when the function only takes finite values) does not follow from the assumption that only the Lebesgue sets of *one certain type* are measurable. What has been said can be confirmed by the following simple example: for the function defined on a nonmeasurable set E and taking the value $-\infty$ at all the points of E all the Lebesgue sets of the first type are empty and thus measurable.

Let us agree that in what follows the functions in question are allowed to assume both finite and infinite values.

Theorem VI.3.2. *If f_k ($k = 1, 2, \dots$) and f are measurable on a set $E \subset X$ the subset $E[f_k(x) \rightarrow f(x)]$ is measurable.*

Proof. Consider the two functions

$$g(x) = \overline{\lim} f_k(x) \quad \text{and} \quad h(x) = \underline{\lim} f_k(x)$$

* Under this assumption, in the definition of a Lebesgue set we can, as before, consider a as an arbitrary real number, and it is unnecessary to admit the values $+\infty$ and $-\infty$ for a . It nevertheless turns out that if f is measurable on E both sets $E[f(x) = +\infty]$ and $E[f(x) = -\infty]$ are measurable. This can be shown, for instance, in the case of the set $E[f(x) = +\infty]$, with the aid of the obvious equality

$$E[f(x) = +\infty] = \bigcap_{n=1}^{\infty} E[f(x) > n]$$

defined for all $x \in E$. Both functions are measurable on E (see the proof of Theorem VI.3.4). The set E' of the points $x \in E$ for which $\lim f_h(x)$ (finite or equal to $\pm\infty$) exists coincides with $E[g(x) = h(x)]$ and therefore is measurable. The set $E'' = E[g(x) = f(x)]$ is also measurable and therefore the set $E[f_h(x) \rightarrow f(x)] = E' \cap E''$ is measurable as well.

§ 4. Equivalent Functions.

Convergence Almost Everywhere

In this section we shall present some additional material concerning the passage to the limit in the class of measurable functions. To this end we state the following

Definition. Let a certain property hold for some points of a set $E \subset X$ (where X is a measure space with measure μ) and not hold for its other points. If the points of E for which the property does not hold are contained in a subset of measure zero this property is said to hold *almost everywhere* (or μ -almost everywhere) on the set E (or for almost all x belonging to E).

For instance, if $f_h(x) \rightarrow f(x)$ for all $x \in E$ except possibly at the points of a subset $E' \subset E$ where $E' \subset E_0$ and $\mu E_0 = 0$ the sequence $\{f_h\}$ is said to converge to f almost everywhere on E .

If a function f is defined on a set $E \subset \mathbf{R}_n$ and the set of all its points of discontinuity is of measure zero we say that the function f is continuous almost everywhere on the set E .

Among the functions which may assume both finite and infinite values those taking the values $+\infty$ and $-\infty$ only on a set of measure zero (or on a subset of such a set) are of particular interest. In this case we speak of functions *finite almost everywhere*.

The property that a finite or countable union of sets of measure zero is also of measure zero readily implies that if we deal with a finite or countable number of some properties each of which holds almost everywhere on a set E all these properties also hold simultaneously almost everywhere on E . For instance, if we are given a sequence of functions $\{f_h\}$

each of which is positive almost everywhere on E the inequality $f_k(x) > 0$ also holds for all $k = 1, 2, \dots$ simultaneously almost everywhere on E .

If the measure μ defined in X is complete the foregoing definition is simplified. Namely, in this case "almost everywhere on E " means "for all $x \in E$ except possibly on a subset $E' \subset E$ with $\mu E' = 0$ ".

As will be seen later, in the general theory of the integral it is often allowable not to take into account some peculiarities of the function in question which only pertain to a point set of measure zero. In this connection we introduce one more

Definition. Two functions f and g defined on a set $E \subset X$ are said to be *equivalent on that set* (written $f \sim g$) if $f(x) = g(x)$ almost everywhere on E .

It is readily seen that this equivalence relation is transitive: if $f \sim g$ on E and $g \sim h$ on the same set then $f \sim h$ on E .

If a measurable function which is finite almost everywhere is redefined by replacing its infinite values by a finite constant the resultant function will only assume finite values and will be measurable and equivalent to the original function.

Theorem VI.4.1. *Let the measure μ in X be complete. Then if two functions f and g are equivalent on a set $E \subset X$ and one of them is measurable on that set the other is also measurable.*

Proof. Let f be measurable on E . Putting

$$E' = E [f(x) = g(x)]$$

and

$$E'' = E [f(x) \neq g(x)]$$

we see that, by the hypothesis, $\mu E'' = 0$ and therefore the set $E' = E \setminus E''$ is measurable. Consequently f and, together with it, g are measurable on E' . Furthermore, g is measurable on E'' since $\mu E'' = 0$ and consequently g is also measurable on E .

It should be noted that this theorem no longer holds if the requirement that the measure should be complete is rejected. For example, let $E \subset X$, $\mu E = 0$ and let $E' \subset E$ be

a nonmeasurable subset. Then the function

$$f(x) = \begin{cases} 1 & \text{for } x \in E' \\ 0 & \text{for } x \in E \setminus E' \end{cases}$$

is obviously equivalent to the measurable function $g(x) \equiv 0$ on E but f is not itself measurable.

Now we shall show that, under the assumption that the measure is complete, Theorem VI.3.1 on the passage to the limit continues to hold not only for the pointwise convergence at all the points x belonging to a measurable set E on which the functions in question are defined but also for the convergence almost everywhere on E .^{*} In particular, this important property holds for the Lebesgue measurable functions in a Euclidean space.

Theorem VI.4.2. *Let the measure μ in X be complete. If f_k ($k = 1, 2, \dots$) and f are defined on a set $E \subset X$, f_k are measurable on E for all k and $f_k(x) \rightarrow f(x)$ almost everywhere on E , the function f is also measurable on E .*

Proof. Let E' be the collection of all points belonging to E at which $f_k(x) \rightarrow f(x)$ and let $E'' = E \setminus E'$. Then E'' is measurable and $\mu E'' = 0$. Since E is measurable** the set $E' = E \setminus E''$ is also measurable and, according to proposition 2° in Sec. VI.1, the functions f_k are measurable on E' . Consequently, by Theorem VI.3.1, f is also measurable on E' . Moreover, f is measurable on E'' because $\mu E'' = 0$ (see proposition 6° in Sec. VI.1), and therefore f is measurable on E as well (see proposition 3° in Sec. VI.1).

Remark. This theorem also fails to hold if the condition of completeness of the measure is rejected. But in this case the following assertion holds: if f_k ($k = 1, 2, \dots$) are measurable functions on a set E , f is a function defined on E and $f_k(x) \rightarrow f(x)$ almost everywhere on E there exists a measurable function g defined on E such that $f_k(x) \rightarrow g(x)$ almost everywhere on E and $g \sim f$. Indeed, there is a (measurable) set $H \subset E$ (with $\mu H = 0$) such that

$$f(x) = \lim f_k(x) \quad \text{for all } x \in E' = E \setminus H$$

^{*} For the convergence almost everywhere (in the case of a complete measure) the corollary of Theorem VI.3.1 also continues to hold.

^{**} Recall that the functions f_k are measurable on E .

Therefore f is measurable on E' . Let us put

$$g(x) = \begin{cases} f(x) & \text{for } x \in E' \\ 0 & \text{for } x \in H \end{cases}$$

It is clear that g satisfies all the required conditions.

In conclusion we note that *without the requirement that the measure should be complete* the following property holds: if f_k and f are measurable on E and $f_k(x) \rightarrow f(x)$ almost everywhere on E the set

$$E[f_k(x) \not\rightarrow f(x)]$$

is measurable and is of measure zero.

This assertion readily follows from Theorem VI.3.2.

§ 5. Convergence in Measure

In the theory of measurable functions an important role is played by a concept of convergence, different from the pointwise convergence, which will be studied in this section. For the sake of simplicity we shall again deal with functions assuming finite values but the reader can easily extend all the facts presented below to the case of measurable functions finite almost everywhere.

Definition. Let f_k ($k = 1, 2, \dots$) and f be measurable on a set $E \subset X$. The sequence $\{f_k\}$ is said to *converge in measure to the function f (on the set E)* if

$$\mu E[|f_k(x) - f(x)| \geq \varepsilon] \xrightarrow[k \rightarrow \infty]{} 0$$

for every $\varepsilon > 0$.*

In the case when $\mu E < +\infty$ this condition means that the measure of the subset of points of E at which $|f_k(x) - f(x)| < \varepsilon$ becomes arbitrarily close to the measure of the whole set E as $k \rightarrow \infty$.**

* The measurability of these sets follows from the results established in Sec. VI.2. In this connection it is important that not only f_k are measurable but also the function f is at the very beginning supposed to be measurable.

** The definition of convergence in measure also applies to functions which are finite almost everywhere. But in this case it should be taken into account that the set $E[|f_k(x) - f(x)| < \varepsilon]$ may not be the complement of $E[|f_k(x) - f(x)| \geq \varepsilon]$. The matter is that the set E may contain a nonempty subset of measure zero, not belonging to the above two sets, at whose points the difference $f_k(x) - f(x)$ does not make sense.

For the convergence in measure we shall use the notation

$$f_h \Rightarrow f$$

A given functional sequence can be convergent in measure to different limit functions. For instance, if $f_h \Rightarrow f$ and $f \sim g$ where g is a measurable function then $f_h \Rightarrow g$. The converse is also true in the sense that any two functions serving as limits in measure of one and the same sequence of measurable functions coincide to within their values on a set of measure zero:

Theorem VI.5.1. *If $f_h \Rightarrow f$ and $f_h \Rightarrow g$ on E then $f \sim g$ on E .*

Before proving the theorem we shall establish the following auxiliary proposition: if $f(x) = \varphi(x) + \psi(x)$ on E then

$$E[|f(x)| \geq \varepsilon] \subset E\left[|\varphi(x)| \geq \frac{\varepsilon}{2}\right] \cup E\left[|\psi(x)| \geq \frac{\varepsilon}{2}\right] \quad (5)$$

for any $\varepsilon > 0$. Indeed, if $x \in E$ and x does not belong to the right-hand member of (5) we have $|\varphi(x)| < \frac{\varepsilon}{2}$ and $|\psi(x)| < \frac{\varepsilon}{2}$ whence $|f(x)| < \varepsilon$, i.e. x is not contained in the left-hand member of (5).

The proof of the theorem. Let $E' = E[f(x) \neq g(x)]$.

Then $E' = \bigcup_{p=1}^{\infty} E_p$ where

$$E_p = E\left[|f(x) - g(x)| \geq \frac{1}{p}\right] \quad (p = 1, 2, \dots)$$

Let us put

$$E'_{pk} = E\left[|f_k(x) - f(x)| \geq \frac{1}{2p}\right]$$

and

$$E''_{pk} = E\left[|f_k(x) - g(x)| \geq \frac{1}{2p}\right]$$

where $p, k = 1, 2, \dots$. By formula (5), $E_p \subset E'_{pk} \cup E''_{pk}$. Consequently,

$$\mu E_p \leq \mu E'_{pk} + \mu E''_{pk} \xrightarrow{k \rightarrow \infty} 0$$

and therefore $\mu E_p = 0$ for all p . It follows that $\mu E'$ is also equal to zero, that is $f \sim g$.

Now we proceed to elucidate the relationship between the convergence in measure and the convergence almost everywhere.

Theorem VI.5.2 (H. Lebesgue). *If a sequence of functions f_k measurable on a set E with $\mu E < +\infty$ is convergent almost everywhere on that set to a measurable function f then $f_k \Rightarrow f$.**

Proof. Let us fix an arbitrary $\varepsilon > 0$ and put

$$A_k = E [|f_k(x) - f(x)| \geq \varepsilon] \quad (k = 1, 2, \dots)$$

Let E' be the subset of all points belonging to E at which the sequence $\{f_k(x)\}$ does not converge to $f(x)$. By the hypothesis, E' is contained in a subset of measure zero (more precisely, as was mentioned at the end of Sec. VI.4, $\mu E' = 0$).

Let us consider the sets $B_k = \bigcup_{p=k}^{\infty} A_p$ and $B = \bigcap_{k=1}^{\infty} B_k$. The set B is obviously measurable. We shall prove that $B \subset E'$.

Indeed, if $x \in B$ then $x \in B_k$ for all k . Therefore for any k there is $p \geq k$ such that $x \in A_p$. Thus, there exist arbitrarily large values of the index p for which

$$|f_p(x) - f(x)| \geq \varepsilon$$

whence $f_k(x) \not\rightarrow f(x)$, that is $x \in E'$. This shows that $B \subset E'$, and hence $\mu B = 0$.

Since the sets B_k form a decreasing sequence with $\mu B_k \leq \mu E < +\infty$, we have $\mu B_k \rightarrow \mu B$, i.e. $\mu B_k \rightarrow 0$. But $A_k \subset B_k$ and consequently $\mu A_k \leq \mu B_k$, which implies $\mu A_k \rightarrow 0$. The theorem has been proved.

Remark. We have in fact proved a stronger assertion: if $f_k(x) \rightarrow f(x)$ almost everywhere on E , $\mu E < +\infty$ (f_k and f are measurable) and

$$B_k = \bigcup_{p=k}^{\infty} E [|f_p(x) - f(x)| \geq \varepsilon]$$

where $\varepsilon > 0$ then $\mu B_k \rightarrow 0$.

* If the function $f(x)$ is regarded as defined by the limiting relation $f(x) = \lim f_k(x)$ at all points x where this limit exists and is finite it can be arbitrarily defined at the remaining points of E without violating its measurability; hence we can assume that f is defined throughout E .

This property will be used in the next section.

The converse of Lebesgue's theorem is not true: a functional sequence can be convergent in measure and, at the same time, can have no limit at every point. To illustrate what has been said we shall consider the following example for the space $\mathbf{R}_1$. Let E be the closed interval $[0, 1]$. Breaking up $E = [0, 1]$ into n equal parts we define the n functions

$$\varphi_n^{(m)}(x) = \begin{cases} 1 & \text{for } \frac{m-1}{n} \leq x < \frac{m}{n} \\ 0 & \text{for all the other } x \in [0, 1] \end{cases}$$

$$(m = 1, 2, \dots, n)$$

Making n assume, in succession, the values $1, 2, \dots$, arranging all the functions $\varphi_n^{(m)}$ as one sequence

$$\varphi_1^{(1)}, \varphi_2^{(1)}, \varphi_2^{(2)}, \varphi_3^{(1)}, \varphi_3^{(2)}, \varphi_3^{(3)}, \dots \quad (6)$$

and putting $f(x) \equiv 0$ on the interval $[0, 1]$ we see that

$$\mu E [\varphi_n^{(m)}(x) \neq f(x)] = \frac{1}{n}$$

Therefore it is clear that sequence (6) is convergent in measure to f . At the same time, for any $x \in [0, 1]$ sequence (6) contains an infinite number of functions assuming the value 1 at that point and an infinite number of functions assuming the value 0 at that point. Consequently, sequence (6) has no limit at every point.

The condition $\mu E < +\infty$ is essential in Theorem VI.5.2 and cannot be discarded, that is the convergence of a functional sequence almost everywhere on a set E with $\mu E = +\infty$ does not, generally, imply the convergence in measure. This is confirmed by the following simple example. Let f_k ($k = 1, 2, \dots$) be the functions defined on the interval $E = (0, +\infty) \subset \mathbf{R}_1$ by means of the relations

$$f_k(x) = \begin{cases} 0 & \text{for } 0 < x \leq k \\ 1 & \text{for } x > k \end{cases}$$

Then $f_k(x) \rightarrow 0$ for all $x \in (0, +\infty)$ but $\mu E [f_k(x) = 1] = +\infty$ and therefore the sequence $\{f_k\}$ is not convergent in measure to the function $f(x) \equiv 0$.

Lemma VI.5.1. *If a sequence of functions f_k measurable on a set E is such that, given any $\varepsilon > 0$, the relation*

$$\mu E[|f_k(x) - f_l(x)| \geq \varepsilon] \xrightarrow[k, l \rightarrow \infty]{} 0 \quad (7)$$

holds, there is a subsequence $\{f_{k_i}\}$ of the given functional sequence which is convergent almost everywhere on E to a measurable (on E) limit function with finite values.

Proof. Let us set positive numbers ε_i so that $\sum_{i=1}^{\infty} \varepsilon_i < +\infty$

and put $\delta_p = \sum_{i=p}^{\infty} \varepsilon_i$ ($p = 1, 2, \dots$). For each i there is a natural number k_i such that

$$\mu E[|f_k(x) - f_l(x)| \geq \varepsilon_i] < \varepsilon_i \quad \text{for } k, l \geq k_i$$

We can assume that $k_1 < k_2 < \dots < k_i < \dots$. Then, in particular,

$$\mu E[|f_{k_{i+1}}(x) - f_{k_i}(x)| \geq \varepsilon_i] < \varepsilon_i$$

Let us put

$$E_p = \bigcup_{i=p}^{\infty} E[|f_{k_{i+1}}(x) - f_{k_i}(x)| \geq \varepsilon_i]$$

Then $\mu E_p < \delta_p$. Lastly, for the intersection $E' = \bigcap_{p=1}^{\infty} E_p$, we obviously have $\mu E' = 0$ (since $\delta_p \rightarrow 0$).

We shall show that the sequence $\{f_{k_i}(x)\}$ has a finite limit $f(x)$ at every point $x \in E \setminus E'$. If $x \in E \setminus E'$ then $x \notin E_p$ for some p . Consequently,

$$|f_{k_{i+1}}(x) - f_{k_i}(x)| < \varepsilon_i \quad \text{for all } i \geq p$$

It follows that for any two indices i and j such that $j > i \geq p$ we have

$$\begin{aligned} |f_{k_j}(x) - f_{k_i}(x)| &\leq \sum_{m=i}^{j-1} |f_{k_{m+1}}(x) - f_{k_m}(x)| < \\ &< \sum_{m=i}^{j-1} \varepsilon_m < \sum_{m=i}^{\infty} \varepsilon_m = \delta_i \end{aligned}$$

Thus,

$$\lim_{i, j \rightarrow \infty} [f_{k_j}(x) - f_{k_i}(x)] = 0$$

This means that $\{f_k(x)\}$ is a Cauchy sequence for every $x \in E \setminus E'$, which implies the pointwise convergence of $\{f_{k_i}\}$ to a finite limit function on the set $E \setminus E'$. Now it only remains to put

$$f(x) = \begin{cases} \lim f_{k_i}(x) & \text{for } x \in E \setminus E' \\ 0 & \text{for } x \in E' \end{cases}$$

to obtain the desired limit function.

Lemma VI.5.2. *If $f_k \Rightarrow f$ on E and, simultaneously, $f_k(x) \rightarrow g(x)$ almost everywhere on E then $f \sim g$ on E .*

Proof. Without losing generality we can assume that the function g is also measurable (see the remark after Theorem VI.4.2). As in the proof of the foregoing lemma, let us

set positive numbers ε_i so that $\sum_{i=1}^{\infty} \varepsilon_i < +\infty$ and then choose an increasing sequence of natural numbers $k_1 < k_2 < \dots < k_i < \dots$ such that

$$\mu E \{ |f_{k_i}(x) - f(x)| \geq \varepsilon_i \} < \varepsilon_i$$

Putting

$$A = \bigcup_{i=1}^{\infty} E \{ |f_{k_i}(x) - f(x)| \geq \varepsilon_i \}$$

we see that if $x \notin A$ then $|f_{k_i}(x) - f(x)| < \varepsilon_i$ for all i ; on passing to the limit in this inequality (we remind the reader that $f_k(x) \rightarrow g(x)$ almost everywhere) we obtain $g(x) = f(x)$ for almost all $x \notin A$. On the other hand, since $\mu A < +\infty$ Theorem VI.5.2 indicates that $f_k \Rightarrow g$ on A and therefore, by Theorem VI.5.1, $f \sim g$ on A . Now it follows that $f \sim g$ on the set E as well.

Theorem VI.5.3 (F. Riesz*). *If $f_k \Rightarrow f$ on a set E there exists a subsequence $\{f_{k_i}\}$ convergent to f almost everywhere on E .*

* F. Riesz (1880-1956) was a distinguished Hungarian mathematician whose most famous results were achieved in the field of functional analysis. Simplifying slightly the proof of Lemma VI.5.1 we can elaborate a direct proof of Theorem VI.5.3. In this section we first prove Lemma VI.5.2 and then obtain the proof of Theorem VI.5.3 because the former is needed for our further aims.

Proof. By virtue of (5), we have

$$\begin{aligned} E[|f_k(x) - f_l(x)| \geq \varepsilon] &\subset \\ &\subset E\left[|f_k(x) - f(x)| \geq \frac{\varepsilon}{2}\right] \cup E\left[|f_l(x) - f(x)| \geq \frac{\varepsilon}{2}\right] \end{aligned}$$

for any k and l whence it follows that the sequence $\{f_k\}$ satisfies condition (7). Therefore, by Lemma VI.5.1, there is a subsequence $\{f_{k_i}\}$ convergent to a function g almost everywhere on E . Since $f_{k_i} \Rightarrow f$, Lemma VI.5.2 implies $f \sim g$. Thus, $f_{k_i}(x) \rightarrow g(x)$ almost everywhere, $g(x) = f(x)$ almost everywhere and, consequently, $f_{k_i}(x) \rightarrow f(x)$ almost everywhere on E .

§ 6. Convergence Regulator. Theorems of D.F. Egorov, N. N. Luzin and R. M. Fréchet

In this section we shall study some further properties of measurable functions and, in particular, prove three fundamental theorems of the theory of measurable functions suggested by D. F. Egorov, N. N. Luzin and R. M. Fréchet*. The first of these theorems will not be used in our course. Before proceeding to these important theorems we shall prove some auxiliary theorems which, by the way, are themselves of interest.

Theorem VI.6.1 (On Stability of Convergence). *If f_k ($k = 1, 2, \dots$) are measurable functions on a set $E \subset X$ with $\mu E < +\infty$ which are finite almost everywhere on E and if $f_k(x) \rightarrow 0$ almost everywhere on E there is an increasing sequence of positive numbers $\lambda_k \rightarrow +\infty$ such that $\lambda_k f_k(x) \rightarrow 0$ almost everywhere on E .*

In a certain sense this property expresses the "stability" of convergence almost everywhere.

Proof. We can suppose, without losing generality, that all the values of the functions f_k are finite (see Sec. VI.4). Furthermore, removing, if necessary, a subset of measure zero from E we can assume that $f_k(x) \rightarrow 0$ for all $x \in E$.

* D. F. Egorov (1869-1931), a Russian mathematician; N. N. Luzin (1883-1950), a noted Soviet mathematician; R. M. Fréchet, a French mathematician (born 1878).

If $f_k(x) \rightarrow 0$ then also $|f_k(x)| \rightarrow 0$. Applying formula (4) to the functions $|f_k(x)|$ we see that there are measurable functions φ_k with finite values forming a decreasing sequence and such that

$$|f_k(x)| \leq \varphi_k(x), \quad \varphi_k(x) \rightarrow 0$$

for all $x \in E$. By Theorem VI.5.2, for every natural number l it is possible to choose k_l so that $k_{l+1} > k_l$ and

$$\mu E \left[\varphi_{k_l}(x) \geq \frac{1}{l^2} \right] < \frac{1}{2^l}$$

Next we put

$$\lambda_k = \begin{cases} 1 & \text{for } k < k_2, \\ l & \text{for } k_l \leq k < k_{l+1} \end{cases} \quad (l = 2, 3, \dots)$$

Let us construct the sets

$$A_l = \bigcup_{k=k_l}^{\infty} E \left[\lambda_k \varphi_k(x) \geq \frac{1}{l} \right] \quad \text{and} \quad B_l = \bigcup_{m=l}^{\infty} A_m$$

The sets A_l can be represented as

$$A_l = \bigcup_{n=l}^{\infty} \bigcup_{k=k_n}^{k_{n+1}-1} E \left[\varphi_k(x) \geq \frac{1}{ln} \right]$$

For $n \geq l$ we have

$$E \left[\varphi_{k_{n+1}}(x) \geq \frac{1}{ln} \right] \subset E \left[\varphi_k(x) \geq \frac{1}{ln} \right] \subset E \left[\varphi_k(x) \geq \frac{1}{n^2} \right]$$

and therefore

$$A_l \subset \bigcup_{n=l}^{\infty} E \left[\varphi_{k_n}(x) \geq \frac{1}{n^2} \right]$$

Consequently,

$$\mu A_l < \sum_{n=l}^{\infty} \frac{1}{2^n} = \frac{1}{2^{l-1}} \quad \text{and} \quad \mu B_l < \sum_{m=l}^{\infty} \frac{1}{2^{m-1}} = \frac{1}{2^{l-2}}$$

Now let us put $A = \bigcap_{l=1}^{\infty} B_l$. Then $\mu A = 0$. We shall show that $\lambda_k \varphi_k(x) \rightarrow 0$ for all $x \notin A$.

Indeed, if $x \notin A$ then $x \notin B_{l_0}$ for some l_0 . Hence, $x \notin A_l$ for all $l \geq l_0$, and therefore $\lambda_k \varphi_k(x) < \frac{1}{l}$ for $k \geq k_l$. It now

becomes clear that $\lambda_k f_k(x) \rightarrow 0$ for all $x \in A$, which is what we set out to prove.

Theorem VI.6.2 (On Convergence Regulator). *If f_k ($k = 1, 2, \dots$) are measurable and finite almost everywhere on a set $E \subset X$ with $\mu E < +\infty$ and if $f_k(x) \rightarrow 0$ almost everywhere on E there exists a measurable nonnegative function g finite almost everywhere on E and a sequence of positive numbers $\alpha_k \rightarrow 0$ such that*

$$|f_k(x)| \leq \alpha_k g(x) \quad \text{for almost all } x \in E$$

This theorem shows that convergence almost everywhere turns out to be "uniform" with respect to some function g . We call such a function g a *convergence regulator* of the sequence $\{f_k\}$.

Proof. By the preceding theorem, it is possible to choose positive numbers $\lambda_k \rightarrow +\infty$ so that $\lambda_k f_k(x) \rightarrow 0$ almost everywhere on E . Let us put

$$g(x) = \sup \lambda_k |f_k(x)|$$

The function g is measurable (see Lemma VI.3.1 which continues to hold for functions with infinite values as well) and finite almost everywhere and $|f_k(x)| \leq \frac{1}{\lambda_k} g(x)$.

In the next chapter we shall present a generalization of the theorem on a convergence regulator.

Theorem VI.6.3 (D. F. Egorov). *Let $\{f_k\}$ be a sequence of measurable functions finite almost everywhere on a set $E \subset X$ (with $\mu E < +\infty$) converging to a function f almost everywhere on E ($f_k(x) \rightarrow f(x)$), the function f also being finite almost everywhere. Then, given any $\varepsilon > 0$, there exists a measurable set $E' \subset E$ such that $\mu E' < \varepsilon$ and $\{f_k\}$ converges to f uniformly on the set $E \setminus E'$.*

This theorem shows that for measurable functions finite almost everywhere convergence almost everywhere "does not essentially differ" from uniform convergence.

Proof. We shall suppose, without loss of generality, that all the values of the functions f_k and f are finite and that the function f is measurable. Let g be the function indicated in Theorem VI.6.2 for the sequence $\{f_k - f\}$ (convergent

to 0). We can write the obvious equality

$$E[g(x) = +\infty] = \bigcap_{p=1}^{\infty} E_p \text{ where } E_p = E[g(x) > p]$$

Since the sets E_p form a decreasing sequence and the function g is finite almost everywhere we have $\mu E_p \rightarrow 0$. Consequently, given an arbitrary $\varepsilon > 0$, there is p such that $\mu E_p < \varepsilon$. Let us put $E' = E_p$. Then, according to the definition of a convergence regulator,

$$|f_k(x) - f(x)| \leq \alpha_k g(x) \leq \alpha_k p$$

for the points of the set $E \setminus E'$, the sequence of positive numbers α_k converging to 0. It follows that $f_k(x) \rightarrow f(x)$ uniformly on $E \setminus E'$.

All the further results of this section will be established for the special case of functions defined in the Euclidean space R_n of an arbitrary dimension n and the letter μ will denote the Lebesgue measure in R_n .

Theorem VI.6.4 (N. N. Luzin). *If f is a measurable function finite almost everywhere on a set $E \subset R_n$ then, given any $\varepsilon > 0$, there is a function φ continuous throughout R_n such that*

$$\mu E[f(x) \neq \varphi(x)] < \varepsilon$$

Proof. To begin with, we shall show that for an arbitrary $\varepsilon > 0$ there exists a closed set $F \subset E$ such that $\mu(E \setminus F) < \varepsilon$ and the restriction of the function f to the set F is continuous on that set.*

(a) We shall first assume that f is bounded on the set E , i.e. $|f(x)| \leq M$. Let us take an arbitrary $\varepsilon > 0$. Next we divide, for any natural k , the interval $[-M, M]$ into $2k$ equal parts by means of the points of division

$$y_i = -M + \frac{iM}{k} \quad (i = 0, 1, 2, \dots, 2k)$$

and construct the sets

$$A_{k1} = E[y_0 \leq f(x) \leq y_1]$$

* This means that if $x_m \in F$ and $x_m \rightarrow x_0$ ($x_0 \in F$) then $f(x_m) \rightarrow f(x_0)$. But at the same time, when $x_m \in E \setminus F$ and $x_m \rightarrow x_0 \in F$, it is possible that $f(x_m) \not\rightarrow f(x_0)$.

and

$$A_{ki} = E [y_{i-1} < f(x) \leq y_i] \quad \text{for } i = 2, 3, \dots, 2k$$

These sets are measurable and disjoint and

$$E = \bigcup_{i=1}^{2k} A_{ki}$$

As follows from Corollary 1 of Theorem V.5.5, for every A_{ki} there exists a closed set $F_{ki} \subset A_{ki}$ such that

$$\mu(A_{ki} \setminus F_{ki}) < \frac{\varepsilon}{2^{k+1}k}$$

Let $F_k = \bigcup_{i=1}^{2k} F_{ki}$. Then

$$E \setminus F_k = \bigcup_{i=1}^{2k} (A_{ki} \setminus F_{ki})$$

and therefore

$$\mu(E \setminus F_k) < \frac{\varepsilon}{2^k} \quad (8)$$

Now we define on F_k the function f_k by putting $f_k(x) = y_i$ for $x \in F_{ki}$ ($i = 1, 2, \dots, 2k$). According to the corollary drawn from Theorem II.7.1, f_k is continuous on F_k . The definition of the sets A_{ki} implies that

$$0 \leq f_k(x) - f(x) < \frac{M}{k} \quad (9)$$

for all $x \in F_k$.

Let us put $F = \bigcap_{k=1}^{\infty} F_k$. Then

$$E \setminus F = \bigcup_{k=1}^{\infty} (E \setminus F_k)$$

and, by (8), $\mu(E \setminus F) < \varepsilon$. Moreover, for $x \in F$ inequality (9) holds for any k , and therefore $f_k(x) \rightarrow f(x)$ uniformly on F . Hence, since all the functions f_k are continuous on F the function f is also continuous on F .

(b) Now we allow the function f to be unbounded but suppose that it is finite almost everywhere on the set E . Let us take the function $g(x) = \arctan f(x)$. If we assume that

$\arctan \pm \infty = \pm \frac{\pi}{2}$ the function $g(x)$ can be considered as defined and bounded throughout E . For any $a \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ we can write

$$E[g(x) > a] = E[\arctan f(x) > a] = E[f(x) > \tan a]$$

the rightmost set being measurable as a Lebesgue set of the function f . If $a \geq \frac{\pi}{2}$ we have $E[g(x) > a] = \emptyset$; if $a = -\frac{\pi}{2}$ then

$$E[g(x) > a] = E[f(x) \neq -\infty] = \bigcup_{n=1}^{\infty} E[f(x) > -n]$$

and if $a < -\frac{\pi}{2}$ then $E[g(x) > a] = E$. Thus, g is measurable, $|g(x)| \leq \frac{\pi}{2}$ throughout E and $|g(x)| < \frac{\pi}{2}$ almost everywhere on E .

Now, by what was proved in (a), it is possible to choose a closed set $F \subset E$ such that $\mu(E \setminus F) < \varepsilon$ and the restriction of the function g to F is continuous. Since we have $f(x) = \tan g(x)$ the restriction of f to F is also continuous.

It can easily be understood that, since f is finite almost everywhere, the set F can be chosen in such a way that f assumes only finite values on it.

(c) To complete the proof of the theorem it remains to show that every function f defined and continuous on a closed set $F \subset \mathbf{R}_n$ can be extended to the entire space $\mathbf{R}_n$ with the retention of the continuity. For the sake of simplicity, in the concluding part of the proof we shall confine ourselves to the case of one-dimensional space $\mathbf{R}_1$.*

Thus, let $F \subset \mathbf{R}_1$ ($F \neq \emptyset$). We know (see Sec. II.4) that there exists the smallest closed interval $[a, b]$ in which the set F can be embedded.** Let us begin with extending f in a continuous fashion from F to that interval $[a, b]$. As was shown in Sec. II.4, F is obtained by deleting from $[a, b]$ a

* For the general case of the space $\mathbf{R}_n$ of arbitrary dimension n we refer the reader to the book by I. P. Natanson, *The Theory of Functions of a Real Variable* (see the bibliography).

** As was agreed (see the second footnote on p. 102), if F is not bounded below (above) we put $a = -\infty$ ($b = +\infty$).

finite or countable system of open intervals. Denoting by $\varphi(x)$ the extension of f to $[a, b]$, which we are going to construct, we put $\varphi(x)$ equal to $f(x)$ for $x \in F$ and define $\varphi(x)$ by means of linear interpolation on each of these open intervals (which are the components of the (open) complement of F with respect to $[a, b]$ (see Fig. 16 where (α, β) is one of these intervals)*). Within each component (α, β) (Fig. 16) the function φ is obviously continuous as a linear function. If $x_0 \in F$ ($x_0 \neq b$) is the left end point of one of the components of $[a, b] \setminus F$ the function φ is continuous

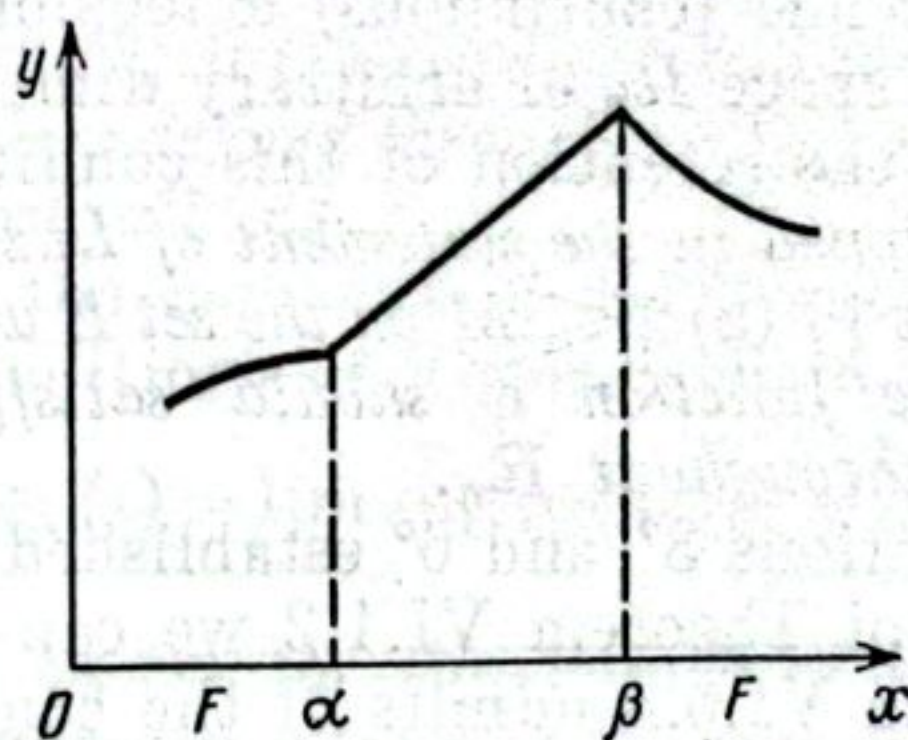


Fig. 16

from the right at the point x_0 also because it is a linear function. Finally, let a point $x_0 \in F$ ($x_0 \neq b$) be different from the left end point of any component of the set $[a, b] \setminus F$. Consider a sequence of points $x_m > x_0$ converging to x_0 , i.e. $x_m \rightarrow x_0 + 0$. If $x_m \in F$ ($m = 1, 2, \dots$) then $\varphi(x_m) = f(x_m) \rightarrow f(x_0) = \varphi(x_0)$ since f is continuous on F . If $x_m \notin F$ ($m = 1, 2, \dots$) then it is quite clear that the end points α_m and β_m of the component (α_m, β_m) of the open set $[a, b] \setminus F$, which contains x_m also approach x_0 (from the right). Therefore in this case we have $f(\alpha_m) \rightarrow f(x_0)$ and $f(\beta_m) \rightarrow f(x_0)$. The value $\varphi(x_m)$ is contained between $f(\alpha_m)$ and $f(\beta_m)$ since φ is linear within every component of $[a, b] \setminus F$, and consequently we have $\varphi(x_m) \rightarrow f(x_0) = \varphi(x_0)$ in this case as well.

It follows that the extended function φ is continuous from the right on the interval $[a, b]$. The continuity of φ from the left on $[a, b]$ is proved similarly.

* The interpolation is not of course needed if $F = [a, b]$.

Now we extend $\varphi(x)$ to the whole space $\mathbb{R}_1$ by putting

$$\varphi(x) = \begin{cases} \varphi(a) & \text{for all } x < a \text{ (if } a > -\infty) \\ \varphi(b) & \text{for all } x > b \text{ (if } b < +\infty) \end{cases}$$

It is apparent that the function φ thus extended to the entire x -axis is continuous everywhere.

Remark. The method of extending a continuous function from a closed set F to the whole real axis used above indicates that if $|f(x)| \leq M$ on F then $|\varphi(x)| \leq M$ for all $x \in (-\infty, \infty)$. The (continuous) extension of a continuous function in the space $\mathbb{R}_n$ of arbitrary dimension can also be employed with the retention of this condition. Thus, *if the function f mentioned in the statement of Luzin's theorem satisfies the condition $|f(x)| \leq M$ on the set E we can additionally require that the function φ should satisfy the inequality $|\varphi(x)| \leq M$ throughout $\mathbb{R}_n$.*

Using propositions 3° and 6° established in Sec. VI.1 and the corollaries of Theorem VI.1.2 we can easily show that Luzin's Theorem VI.6.4 admits of the converse: if the conclusion of this theorem holds for a function f (which is not originally supposed to be measurable) defined on a set $E \subset \mathbb{R}_n$ and finite almost everywhere on E this function is measurable. The exact statement of the converse can be, for instance, read as follows: *if a function f is defined on a measurable set $E \subset \mathbb{R}_n$ and is finite almost everywhere on E and if for any $\varepsilon > 0$ there exists a measurable subset $E' \subset E$ on which f is continuous such that $\mu(E \setminus E') < \varepsilon$ the function f is measurable on E .* Thus, the property of measurable functions established in Theorem VI.6.4 completely characterizes this class of functions. In other words, the structure of measurable functions in $\mathbb{R}_n$ is closely connected with that of continuous functions.

The theorem below suggested by R. M. Fréchet can be obtained as a consequence of Luzin's theorem.

Theorem VI.6.5. *If f is measurable and finite almost everywhere on a set $E \subset \mathbb{R}_n$ there exists a sequence of functions φ_k continuous throughout $\mathbb{R}_n$ such that $\varphi_k(x) \rightarrow f(x)$ almost everywhere on E .*

Proof. Let us set a sequence of positive numbers $\varepsilon_i \rightarrow 0$. By Luzin's theorem, for every i there is a continuous fun-

ction g_i such that

$$\mu E [f(x) \neq g_i(x)] < \varepsilon_i$$

Hence, the sequence $\{g_i\}$ converges in measure to f on the set E . By virtue of Theorem VI.5.3, there exists a subsequence $\{g_{i_k}\}$ such that $g_{i_k}(x) \rightarrow f(x)$ almost everywhere on E . To complete the proof it only remains to put $\varphi_k = g_{i_k}$.

From Theorem VI.6.5 we can deduce the following theorem proved by the Italian mathematician G. Vitali (1875-1932).

Theorem VI.6.6. *If f is measurable and finite almost everywhere on a set $E \subset \mathbf{R}_n$ it is equivalent to a function of at most the second Baire class defined throughout $\mathbf{R}_n$, such that $f(x) = g(x)$ for almost all $x \in E$.*

Proof. Using the sequence of continuous functions φ_k satisfying the conditions of Theorem VI.6.5 we put

$$g(x) = \overline{\lim} \varphi_k(x) = \lim_{k \rightarrow \infty} \sup \{\varphi_k(x), \varphi_{k+1}(x), \dots\}$$

Then $g(x) = \lim \varphi_k(x)$ for the points at which this limit exists, and consequently $g(x) = f(x)$ almost everywhere on E . Since the function

$$\begin{aligned} g_k(x) &= \sup \{\varphi_k(x), \varphi_{k+1}(x), \dots\} = \\ &= \lim_{p \rightarrow \infty} \sup \{\varphi_k(x), \varphi_{k+1}(x), \dots, \varphi_{k+p}(x)\} \end{aligned}$$

is of at most the first Baire class we see that g is of at most the second Baire class.

If the function f is bounded on E , i.e. $|f(x)| \leq M$, then, according to the remark made after the proof of Luzin's theorem, the function g can be constructed in such a way that it also satisfies the condition $|g(x)| \leq M$ throughout the space $\mathbf{R}_n$. From the proof of Vitali's theorem presented above it does not follow that the function g necessarily takes finite values at all the points. But a slight modification of this proof makes it possible to show that the function $g(x)$ can be chosen so that it is finite everywhere.

The Lebesgue Integral of a Bounded Function

§ 1. The Definition of the Lebesgue Integral

In the classical definition of the Riemann integral the value of the integral is defined as the limit of the Riemann integral sums. As is known, all the continuous functions turn out to be Riemann integrable; although some discontinuous functions are also Riemann integrable, the class of such functions is rather narrow. This is accounted for by the fact that in the construction of the Riemann integral the domain of integration is partitioned into sets of a comparatively simple shape. For instance, in the case of one-dimensional space the interval of integration is only split into subintervals and only such a simple concept as the length of an interval is used. The introduction of the concept of measure enabled H. Lebesgue to suggest a new definition of the integral for which the class of integrable functions is considerably wider. In the presentation of the general theory of the integral we shall not follow the original scheme of Lebesgue.

In this chapter we shall consider the Lebesgue integral in an arbitrary abstract measure space X with measure μ for the special case of *bounded* functions, and the integrals in question will only be taken over sets with *finite measure* (the generalization to unbounded functions and to domains of integration of infinite measure will be studied in Chapter VIII). As before, we shall designate by $\mathcal{G}$ the σ -algebra of subsets of X on which the measure μ is defined, the sets belonging to $\mathcal{G}$ being called measurable.

Let a bounded function f be defined on a measurable set $E \subset X$ with $\mu E < +\infty$. Breaking the set E in an arbitrary way into a finite number of disjoint measurable sets

$e_i \left(E = \bigcup_{i=1}^p e_i \right)$ we put

$$M_i = \sup_{x \in e_i} f(x) \quad \text{and} \quad m_i = \inf_{x \in e_i} f(x) \quad (i = 1, 2, \dots, p)^*$$

A partition of E into the sets e_i of that type will itself be denoted by τ . Now we form the two sums

$$S(\tau, f) = \sum_{i=1}^p M_i \mu e_i \quad \text{and} \quad s(\tau, f) = \sum_{i=1}^p m_i \mu e_i$$

which will be called, respectively, the *upper* and the *lower Lebesgue-Darboux** sums* (corresponding to the given partition τ). For these sums we shall also use the shortened notation $S(\tau)$, $s(\tau)$ when the function for which they are formed has already been stipulated.

Let us establish some properties of the Lebesgue-Darboux sums. Consider a partition τ of the set E into sets e_i and another partition τ' of E into sets e'_j . We shall say that *the partition τ' is a refinement of the partition τ* if each e'_j is contained in some e_i ; in other words, the sets e'_j are obtained by means of a further splitting of some (or all) sets e_i .

1°. If a partition τ' is a refinement of a partition τ then $S(\tau') \leq S(\tau)$ and $s(\tau') \geq s(\tau)$.

Proof. Since the passage from τ to τ' can be obtained by means of step-by-step refinement it is sufficient to prove proposition 1° for the case when τ' is obtained from τ by the division of one of the sets e_i , for definiteness, e_1 , into two disjoint parts e'_1 and e''_1 . Putting

$$M'_1 = \sup_{x \in e'_1} f(x) \quad \text{and} \quad M''_1 = \sup_{x \in e''_1} f(x)$$

we have $M'_1 \leq M_1$, $M''_1 \leq M_1$ while $\mu e_1 = \mu e'_1 + \mu e''_1$. Consequently,

$$S(\tau') = M'_1 \mu e'_1 + M''_1 \mu e''_1 + \sum_{i=2}^p M_i \mu e_i \leq$$

$$\leq M_1 (\mu e'_1 + \mu e''_1) + \sum_{i=2}^p M_i \mu e_i = \sum_{i=1}^p M_i \mu e_i = S(\tau)$$

* We of course suppose that the sets e_i are nonempty.

** J. G. Darboux (1842-1917), a French mathematician.

that is the upper Lebesgue-Darboux sum does not increase when we pass from τ to τ' . The fact that the lower sum does not decrease is proved similarly.

2°. Any lower Lebesgue-Darboux sum $s(\tau')$ does not exceed any upper Lebesgue-Darboux sum $S(\tau'')$ even when they are formed for two different partitions τ' and τ'' of the set E .

Proof. For the sums corresponding to one and the same partition the inequality $s(\tau) \leq S(\tau)$ is obvious. Suppose now that τ' and τ'' are two different partitions:

$$\tau' : E = \bigcup_{i=1}^p e'_i \quad \text{and} \quad \tau'' : E = \bigcup_{j=1}^q e''_j$$

Let us pass to the new partition τ of the set E into the sets $e_{ij} = e'_i \cap e''_j$ (omitting those e_{ij} which are empty). Then τ simultaneously serves as a refinement of τ' and as a refinement of τ'' , and, according to proposition 1°,

$$s(\tau) \geq s(\tau') \quad \text{and} \quad S(\tau) \leq S(\tau'')$$

Since $s(\tau) \leq S(\tau)$ we obtain $s(\tau') \leq S(\tau'')$.

It follows from 2° that the set of all lower Lebesgue-Darboux sums $s(\tau)$ corresponding to all the possible partitions τ of the set E is bounded above by any upper sum. Consequently, $K = \sup_{\tau} s(\tau; f) \leq S(\tau; f)$. Thus, K serves as a lower bound for the set of the upper sums, and therefore

$$L = \inf_{\tau} S(\tau; f) \geq K = \sup_{\tau} s(\tau; f)$$

Definition. A function f is said to be (Lebesgue) *integrable** with respect to the given measure μ (or, briefly, μ -integrable) on the set E if $K = L$; in this case the common value of the numbers K and L is called the *Lebesgue integral* of the function f over the set E and is denoted as $\int_E f d\mu$ or (sometimes)

$$\int_E f(x) d\mu.$$

* For a Lebesgue integrable function f the term (Lebesgue) *summable* function is also synonymously used. In this course the term "integrable" is only applied to bounded functions integrated over a set with finite measure while the term "summable" (see Chapter VIII) is used for integrable functions which are unbounded or are integrated over a set with infinite measure (or both).—Tr.

If E is an interval $[a, b]$ belonging to $\mathcal{R}_1$ and μ is the Lebesgue measure the classical notation $\int_a^b f(x) dx$ or $\int_a^b f dx$ is also used.

The following theorem indicates that the class of Lebesgue integrable functions is rather wide.

Theorem VII.1.1. *Every bounded function f measurable on a set E is integrable on E .*

Proof. For any partition τ we have

$$s(\tau) \leq K \leq L \leq S(\tau)$$

Hence, to prove that $K = L$ it is sufficient to show that there are partitions τ of the set E for which the upper and the lower Lebesgue-Darboux sums are arbitrarily close to each other.

Let $A < f(x) < B$ for all $x \in E$ (A and B are finite numbers). Breaking the interval (A, B) into a finite number of subintervals with the aid of points of division $A = \lambda_0 < \lambda_1 < \dots < \lambda_p = B$ we put δ equal to the greatest of the differences between two neighbouring numbers λ_i :

$$\delta = \max (\lambda_i - \lambda_{i-1}) \quad (i = 1, 2, \dots, p)$$

Let us form the sets

$$e_i = E [\lambda_{i-1} \leq f(x) < \lambda_i] \quad (i = 1, 2, \dots, p)$$

Each of the sets e_i is measurable as the intersection of two Lebesgue sets of the function f , the sets e_i are pairwise disjoint and $E = \bigcup_{i=1}^p e_i$ (see Fig. 17)*. It follows, in particular that

$$\sum_{i=1}^p \mu e_i = \mu E$$

* This construction in which we first partition the interval containing the range of the values of the function (in our case, the interval (A, B)) and, proceeding from this partition, form the corresponding partition of the domain of definition of the function was suggested by H. Lebesgue in his scheme of the definition of the integral.

Let τ denote the partition of the set E into these sets e_i (if some of e_i 's are empty we simply omit them). Then, for each i , we have

$$\lambda_{i-1} \leq m_i \leq M_i \leq \lambda_i$$

and therefore $M_i - m_i \leq \delta$ and

$$S(\tau) - s(\tau) = \sum_{i=1}^p (M_i - m_i) \mu e_i \leq \delta \sum_{i=1}^p \mu e_i = \delta \mu E$$

The difference $S(\tau) - s(\tau)$ can thus be made arbitrarily

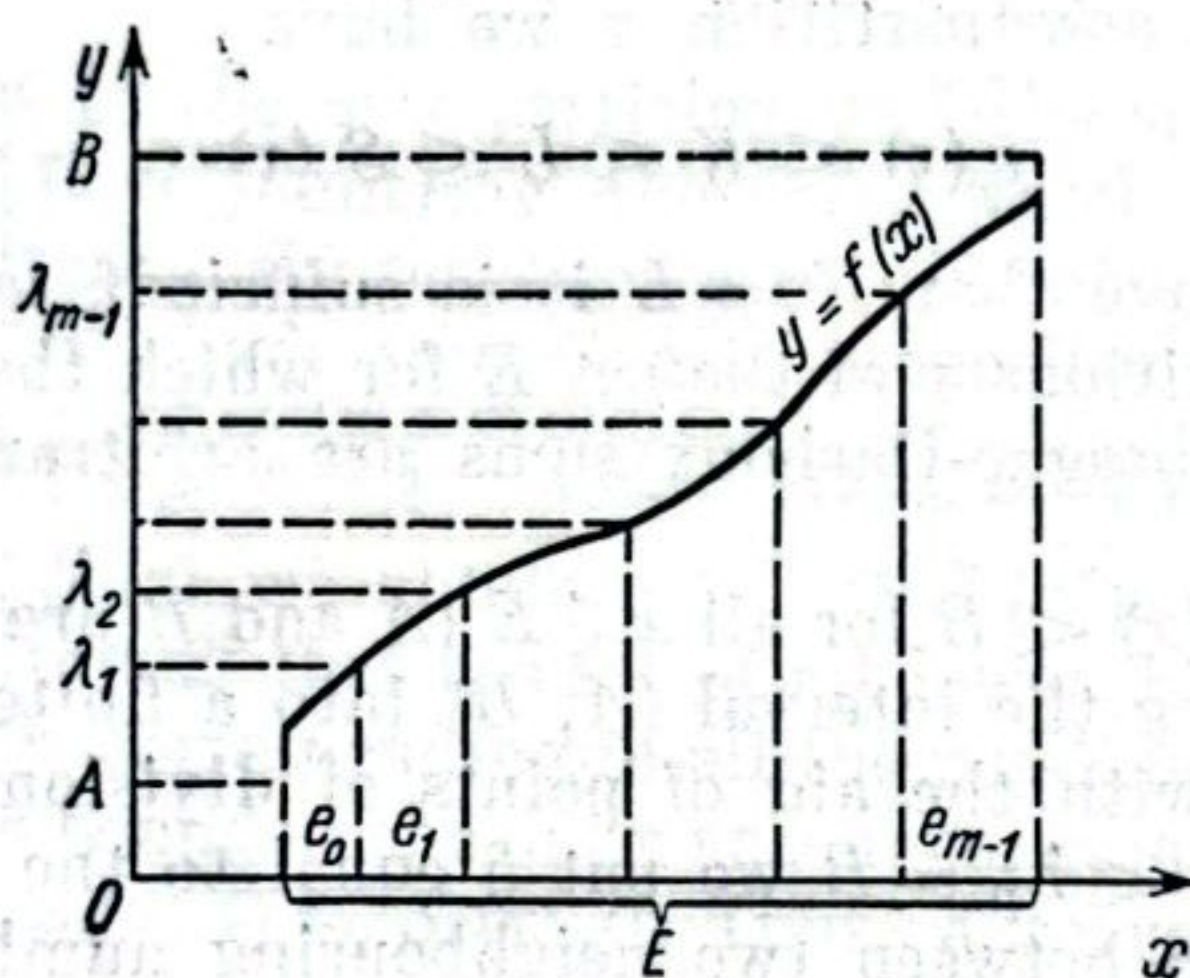


Fig. 17

small (by means of an appropriate choice of the arbitrary quantity δ), which is what we intended to prove.

An important special case of the notion of the integral defined above appears when f is a function in a Euclidean space in which the Lebesgue measure is taken as μ .^{*} From Theorem VI.1.2 and from the above theorem we readily deduce the following corollary for functions in a Euclidean space.

Corollary. *If a function f is continuous on a bounded closed set $E \subset \mathbb{R}_n$ it is Lebesgue integrable on E .*

As is known, the value of the classical (Riemann) integral of a continuous function taken over a bounded closed domain D lies between the values of the Darboux sums. The

^{*} The original scheme was applied by H. Lebesgue to the development of the theory of the integral for the one-dimensional space.

Darboux sums are a special case of the Lebesgue-Darboux sums; the former are constructed on the basis of the same principle as the latter but the domain D is partitioned into sets of a certain particular type, for instance, into subdomains, instead of arbitrary measurable sets. Therefore the value of the Lebesgue integral of a continuous function is sure to lie between those of Darboux's sums. On the other hand, there is a single number lying between the values of all lower and all upper Darboux' sums of a continuous function, namely, the value of the Riemann integral of that function. Hence, *the Lebesgue and the Riemann integrals of a continuous function taken over a bounded closed domain coincide*. The same conclusion can be analogously drawn for an arbitrary bounded Riemann integrable function.

Let us discuss the following two examples.

1. Let X be an arbitrary infinite set containing a countable subset of points $x_1, x_2, \dots, x_k, \dots$ and let there be a convergent positive series $\sum_{k=1}^{\infty} \mu_k$ where $\mu_k > 0$ for all k .

Taking an arbitrary subset $E \subset X$ we put $\mu E = \sum_{k(x_k \in E)} \mu_k$ where the summation extends over all k 's for which $x_k \in E$.* Using the physical terminology we can say that there are *charges* of magnitudes μ_k ($k = 1, 2, \dots$) placed at the points x_k and that μE is equal to the sum of all charges falling inside E .** It is obvious that the set function μ thus defined is a finite measure in X .

If now we take an arbitrary bounded function f defined on a subset E of this space X (it is necessarily measurable since all the subsets of X are measurable) it can be easily verified that the integral of this function is expressed by the formula

$$\int_E f d\mu = \sum_{k(x_k \in E)} f(x_k) \mu_k$$

* If there are no such values of the index k we agree that the sum is equal to zero.

** Since in this example the "point loads" μ_k are positive we can also call them "masses". But the term "charge" is preferable since it can be retained in the general case when such "loads" can be of arbitrary sign (see Sec. XI.1).

We can also denote by $\mu(x)$ the measure of the single-element set (x) for any $x \in E$ and write

$$\int_E f d\mu = \sum_{x \in E} f(x) \mu(x)$$

which means that the integral is equal to the sum of "weighted" values of the functions f taken over all the points of the set E . In this interpretation the values of f at the points not coinciding with any x_k are taken with "weight" equal to 0.

2. Let A be an arbitrary measurable subset of an arbitrary measure space X with measure μ . Taking an arbitrary measurable set $E \subset X$ we put $\nu E = \mu(E \cap A)$. The set function ν is a measure defined on the same σ -algebra of subsets of X on which the measure μ is defined. If f is a bounded function measurable on a set E with $\nu E < +\infty$ then

$$\int_E f d\nu = \int_{E \cap A} f d\mu$$

We also mention without proof that Theorem VII.1.1 admits of a converse: *if a bounded function f is integrable on a measurable set $E \subset X$ with $\mu E < +\infty$ and the measure μ is complete the function f is measurable on E .**

§ 2. The Simplest Properties of the Integral

Now we proceed to establish a number of simple properties of the Lebesgue integral. As a rule, in what follows integrand functions are supposed to be bounded and measurable, and the integrals are taken over sets of finite measure. It should be noted that if $\mu E = 0$ any bounded function f defined on E is integrable on E .** This property is a trivial consequence of the definition of the integral. We shall also agree that the integral of any function taken over an empty set makes sense and is understood as equal to zero.

* Rejecting the requirement that the measure μ should be complete we can prove that if a bounded function is integrable it is equivalent to a bounded measurable function.

** This remark is of interest only when the measure μ is not complete (cf. proposition 6° in Sec. VI.1).

Theorem VII.2.1 (The Estimation of the Integral). *If $C \leq f(x) \leq D$ and f is measurable on a set E then*

$$C\mu E \leq \int_E f d\mu \leq D\mu E \quad (1)$$

Proof. If we take as τ the partition of the set E involving only the set E itself, that is if we put $p=1$ and $e_1 = E$, then $s(\tau) = m \cdot \mu E$ and $S(\tau) = M \cdot \mu E$ where $m = \inf_{x \in E} f(x)$ and $M = \sup_{x \in E} f(x)$. But $C \leq m \leq M \leq D$, and therefore

$$C\mu E \leq s(\tau) \leq \int_E f d\mu \leq S(\tau) \leq D\mu E$$

Corollaries. 1. *If $f(x) \geq 0$ on E then $\int_E f d\mu \geq 0$ (this follows from (1) for $C = 0$).*

2. *If $f(x)$ is equal to a constant p on a set E then $\int_E f d\mu = p \cdot \mu E$ (this is implied by (1) if we take $C = D = p$).*

Theorem VII.2.2 (The σ -additivity of the Integral). *Let E be a set representable as a finite or countable union of pairwise disjoint measurable sets E_j ($E = \bigcup_j E_j$). Then for every function f bounded and measurable on the set E we have*

$$\int_E f d\mu = \sum_j \int_{E_j} f d\mu \quad (2)$$

Proof. We shall first prove that equality (2) holds in the case when E is split into two disjoint subsets: $E = E_1 \cup E_2$ where $E_1 \cap E_2 = \emptyset$.

Let us take an arbitrary partition τ of the set E : $E = \bigcup_{i=1}^p e_i$. Then we put

$$e'_i = e_i \cap E_1, \quad e''_i = e_i \cap E_2 \quad (i = 1, 2, \dots, p)$$

and thus obtain the partitions τ' and τ'' of, respectively, the sets E_1 and E_2 :

$$\tau' : E_1 = \bigcup_i e'_i \quad \text{and} \quad \tau'' : E_2 = \bigcup_i e''_i$$

(those e'_i and e''_i which are empty are discarded). Now we can write

$$s(\tau') \leq \int_{E_1} f d\mu \leq S(\tau'), \quad s(\tau'') \leq \int_{E_2} f d\mu \leq S(\tau'') \quad (3)$$

Combining all the sets e'_i and e''_i we obtain a new partition τ^* of the set E which is a refinement of τ , and

$$s(\tau^*) = s(\tau') + s(\tau''), \quad S(\tau^*) = S(\tau') + S(\tau'')$$

Using the last relations and equality (3) and taking into account proposition 1° established in Sec. VII.1 we find that

$$s(\tau) \leq s(\tau^*) \leq \int_{E_1} f d\mu + \int_{E_2} f d\mu \leq S(\tau^*) \leq S(\tau)$$

Now, since the integral $\int_E f d\mu$ is the only number lying between $s(\tau)$ and $S(\tau)$ for any τ , we conclude that

$$\int_E f d\mu = \int_{E_1} f d\mu + \int_{E_2} f d\mu$$

The next step is to prove by induction that formula (2) also holds when E is broken up into any finite number of disjoint measurable sets, which is quite obvious.

Let us pass to the case of a countable union: $E = \bigcup_{j=1}^{\infty} E_j$.

Putting $H_p = \bigcup_{j=p+1}^{\infty} E_j$ we obtain

$$E = E_1 \cup E_2 \cup \dots \cup E_p \cup H_p \quad (4)$$

where the sets entering into the right-hand side are measurable. Moreover, we have $\mu H_p = \sum_{i=p+1}^{\infty} \mu E_i$, that is μH_p is

equal to the remainder of the convergent series $\sum_{j=1}^{\infty} \mu E_j = \mu E$.

Therefore $\mu H_p \rightarrow 0$.

Since formula (4) involves a finite union we can write, by what has already been proved,

$$\int_E f d\mu = \sum_{j=1}^p \int_{E_j} f d\mu + \int_{H_p} f d\mu$$

From Theorem VII.2.1 it follows that $\int_{H_p} f d\mu \xrightarrow{p \rightarrow \infty} 0$, and therefore

$$\int_E f d\mu = \lim_{p \rightarrow \infty} \sum_{j=1}^p \int_{E_j} f d\mu = \sum_{j=1}^{\infty} \int_{E_j} f d\mu$$

which is what we set out to prove.

The theorem proved indicates that the integral $\int_e f d\mu$ of a given (fixed) function f is a σ -additive set function defined on the σ -algebra $\mathfrak{S}_E$ of all measurable subsets $e \subset E$. If $f(x) \geq 0$ on E we have $\int_e f d\mu \geq 0$ for any $e \in \mathfrak{S}_E$, that is in this case the integral is a (new) measure defined on $\mathfrak{S}_E$.

Corollaries. 1. If $E = \bigcup_{j=1}^p E_j$ where E_j are measurable and pairwise disjoint and $f(x) = c_j$ on E_j ($j = 1, 2, \dots, p$; c_j are constants) then

$$\int_E f d\mu = \sum_{j=1}^p c_j \mu E_j$$

This follows immediately from Theorem VII.2.2 and Corollary 2 of Theorem VII.2.1.

2. If $E = \bigcup_{m=1}^{\infty} E'_m$ where the sets E'_m are measurable and form an increasing sequence then

$$\int_E f d\mu = \lim_{m \rightarrow \infty} \int_{E'_m} f d\mu$$

This is implied by Theorem IV.1.2.

3. If f is bounded, measurable and nonnegative on E ($f(x) \geq 0$) and E' is a measurable subset of E then

$$\int_{E'} f d\mu \leq \int_E f d\mu$$

To prove this proposition we can resort to the interpretation of the integral as a measure and to the property of monotonicity of measure.

Theorem VII.2.3. *For any two functions f and g bounded and measurable on E and any constant c we have*

$$\int_E (f \pm g) d\mu = \int_E f d\mu \pm \int_E g d\mu \quad (5)$$

and

$$\int_E cf d\mu = c \int_E f d\mu^* \quad (6)$$

Proof. We first establish equality (5) for the sum $h = f + g$ of two functions. Let τ be an arbitrary partition of the set E into measurable disjoint subsets e_i , i.e. $E = \bigcup_{i=1}^p e_i$, and let

$$m'_i = \inf_{x \in e_i} f(x), \quad m''_i = \inf_{x \in e_i} g(x), \quad m_i = \inf_{x \in e_i} h(x)$$

$$M'_i = \sup_{x \in e_i} f(x), \quad M''_i = \sup_{x \in e_i} g(x), \quad M_i = \sup_{x \in e_i} h(x)$$

Then

$$m'_i + m''_i \leq h(x) \leq M'_i + M''_i \quad \text{for } x \in e_i$$

and therefore, for any $i = 1, 2, \dots, p$, we have

$$m_i \geq m'_i + m''_i \quad \text{and} \quad M_i \leq M'_i + M''_i$$

Furthermore,

$$\sum_{i=1}^p m'_i \mu e_i \leq \int_E f d\mu \leq \sum_{i=1}^p M'_i \mu e_i$$

and

$$\sum_{i=1}^p m''_i \mu e_i \leq \int_E g d\mu \leq \sum_{i=1}^p M''_i \mu e_i$$

where the leftmost and the rightmost members of the inequalities can be made arbitrarily close to each other by

* The integrals on the left-hand sides of both formulas are sure to exist since the functions $f \pm g$ and cf are bounded and measurable on E .

means of an appropriate choice of the partition τ . Consequently,

$$\sum_{i=1}^p (m'_i + m''_i) \mu e_i \leq \int_E f d\mu + \int_E g d\mu \leq \sum_{i=1}^p (M'_i + M''_i) \mu e_i \quad (7)$$

At the same time,

$$\begin{aligned} \sum_{i=1}^p (m'_i + m''_i) \mu e_i &\leq \sum_{i=1}^p m_i \mu e_i \leq \int_E h d\mu \leq \sum_{i=1}^p M_i \mu e_i \leq \\ &\leq \sum_{i=1}^p (M'_i + M''_i) \mu e_i \end{aligned} \quad (8)$$

Since the rightmost and the leftmost terms of inequalities (7) and (8) are arbitrarily close to each other the middle terms coincide, which proves formula (5) for the sum of two functions.

Now we proceed to prove formula (6).

Let us begin with the case $c \geq 0$. Then

$$\begin{aligned} \int_E cf d\mu &= \sup_{\tau} s(\tau; cf) = \sup_{\tau} c \cdot s(\tau; f) = \\ &= c \sup_{\tau} s(\tau; f) = c \int_E f d\mu^* \end{aligned}$$

In the case when $c < 0$ we write

$$\begin{aligned} \int_E cf d\mu &= \sup_{\tau} s(\tau; cf) = \sup_{\tau} c \cdot S(\tau; f) = \\ &= c \inf_{\tau} S(\tau; f) = c \int_E f d\mu \end{aligned}$$

Finally, formula (5) for the difference of two functions readily follows from what has already been proved.

* Here we make use of the obvious equalities

$$\sup \{ca_{\xi}\} = c \cdot \sup \{a_{\xi}\} \text{ and } \inf \{ca_{\xi}\} = c \cdot \inf \{a_{\xi}\}$$

which hold for any bounded set of numbers a_{ξ} if $c \geq 0$. In particular, it follows that

$$\inf_{x \in e} cf(x) = c \cdot \inf_{x \in e} f(x)$$

and therefore $s(\tau; cf) = c \cdot s(\tau; f)$.

Theorem VII.2.4. *For any bounded measurable function f there holds the inequality*

$$\left| \int_E f d\mu \right| \leq \int_E |f| d\mu^*$$

Proof. Let us put

$$E_1 = E [f(x) \geq 0] \quad \text{and} \quad E_2 = E [f(x) < 0]$$

The sets E_1 and E_2 are measurable, $E_1 \cap E_2 = \emptyset$ and $E_1 \cup E_2 = E$. Hence, by the additivity of the integral,

$$(8) \quad \int_E f d\mu = \int_{E_1} f d\mu + \int_{E_2} f d\mu = \int_{E_1} |f| d\mu - \int_{E_2} |f| d\mu$$

Consequently

$$\left| \int_E f d\mu \right| \leq \int_{E_1} |f| d\mu + \int_{E_2} |f| d\mu = \int_E |f| d\mu$$

Theorem VII.2.5. *If $f \sim g$ on a set E (both functions are supposed to be bounded and measurable) then*

$$\int_E f d\mu = \int_E g d\mu^{**} \quad (9)$$

Proof. Let us put

$$E_1 = E [f(x) \neq g(x)] \quad \text{and} \quad E_2 = E \setminus E_1$$

Since $\mu E_1 = 0$ we have

$$\int_{E_1} f d\mu = \int_{E_1} g d\mu = 0$$

On the other hand, $f(x) = g(x)$ on E_2 and therefore

$$\int_{E_2} f d\mu = \int_{E_2} g d\mu$$

* The measurability of the function $|f|$ was proved in Lemma VI 2.1.

** The following more general proposition can also be easily proved: if f and g are bounded on a set E , $f \sim g$ on E and f is integrable on that set the function g is also integrable on E and equality (9) holds. In particular, if a bounded function is equivalent to a measurable one the former is integrable.

Whence, by the additivity of the integral, we obtain equality (9).

Remark. The theorem proved means that any variation of the values of the integrand function on a set of measure zero does not affect the value of the integral.* Therefore we can also consider the integral $\int_E f d\mu$ of a measurable (and,

in this chapter, also bounded) function f which is defined only almost everywhere on E without specifying the values of the function f on the remaining part of the set E .**

Corollary 1 of Theorem VII.2.1 can be slightly strengthened: if $f(x) \geq 0$ almost everywhere on E (f is bounded and measurable) then $\int_E f d\mu \geq 0$. Indeed, putting

$$g(x) = \begin{cases} f(x) & \text{if } f(x) \geq 0 \\ 0 & \text{if } f(x) < 0 \end{cases}$$

we obtain the measurable function $g(x)$ equivalent to f . Since $g(x) \geq 0$ for all $x \in E$ there must be $\int_E g d\mu \geq 0$.

Consequently, by Theorem VII.2.5, we also have $\int_E f d\mu \geq 0$.

Similarly, if $f(x) = p$ ($p = \text{const}$) almost everywhere on E then $\int_E f d\mu = p \cdot \mu E$. In particular, if $f \sim 0$ (i.e. $f(x) = 0$ almost everywhere) we have $\int_E f d\mu = 0$.

This assertion admits of a converse:

* We remind the reader that in this chapter we only consider integrals of bounded functions, and therefore the variation of the values of the function on a set of measure zero should be such that they do not violate the boundedness of the function. In the next chapter we shall reject this condition.

** Such a function f can always be extended to the remaining part of E with the retention of the measurability property. For example, if f is measurable on $E' \subset E$ and $\mu(E \setminus E') = 0$ we can simply put $f(x) = 0$ on $E \setminus E'$. But the remark made after the proof of Theorem VII.2.5 shows that the requirement that the measurability should be retained is inessential.

Theorem VII.2.6. *If $f(x) \geq 0$ almost everywhere on E (f is supposed to be bounded and measurable) and $\int_E f d\mu = 0$ then $f \sim 0$.*

Proof. Let us put

$$E_1 = E[f(x) > 0], \quad E_2 = E[f(x) < 0] \quad \text{and} \quad E_3 = E[f(x) = 0]$$

By the hypothesis, $\mu E_2 = 0$, and therefore we also have $\int_{E_2} f d\mu = 0$. Furthermore, $\int_{E_3} f d\mu = 0$, and consequently

$$\int_{E_1} f d\mu = \int_E f d\mu = 0 \quad (10)$$

It is clear that

$$E_1 = \bigcup_{m=1}^{\infty} H_m \quad \text{where} \quad H_m = E\left[f(x) \geq \frac{1}{m}\right]$$

The estimation of the integral given in Theorem VII.2.1 shows that

$$\int_{H_m} f d\mu \geq \frac{1}{m} \mu H_m \quad \text{for any } m$$

On the other hand, since $f(x) > 0$ throughout E_1 , we have

$$\int_{E_1} f d\mu \geq \int_{H_m} f d\mu$$

(see Corollary 3 of Theorem VII.2.2) and therefore

$$\int_{E_1} f d\mu \geq \frac{1}{m} \mu H_m$$

whence, by (10), it follows that $\mu H_m = 0$ for all m . Hence, we also have $\mu E_1 = 0$. Thus, $\mu(E_1 \cup E_2) = 0$, which exactly means that $f \sim 0$.

Theorem VII.2.7 (On Termwise Integration of Inequalities). *If $f(x) \leq g(x)$ almost everywhere on a set E (both functions are bounded and measurable) then*

$$\int_E f d\mu \leq \int_E g d\mu$$

Proof. Since $g(x) - f(x) \geq 0$ almost everywhere on E we have

$$\int_E g d\mu - \int_E f d\mu = \int_E (g - f) d\mu \geq 0$$

§ 3. Passage to the Limit in the Lebesgue Integral

Let us consider a set $E \subset X$ with $\mu E < +\infty$ and a sequence of bounded measurable functions f_k ($k = 1, 2, \dots$) defined on E . Suppose that $f_k \Rightarrow f$ where the function f is also bounded (and of course measurable). If for such a sequence the limiting relation

$$\int_E f_k d\mu \rightarrow \int_E f d\mu \quad (11)$$

holds we say that for the given functional sequence on the set E it is allowable to pass to the limit under the integral sign.

Property (11) may not hold even when $f_k(x) \rightarrow f(x)$ at all the points $x \in E$. For instance, taking as E the interval $(0, 1)$ and putting

$$f_k(x) = \begin{cases} k & \text{for } 0 < x < \frac{1}{k} \\ 0 & \text{for } \frac{1}{k} \leq x < 1 \end{cases} \quad (k = 1, 2, \dots)$$

we have $f_k(x) \rightarrow 0$ for all $x \in (0, 1)$, i.e. $f(x) \equiv 0$, while relation (11) does not hold because, by Corollary 1 of Theorem VII.2.2,

$$\int_0^1 f_k d\mu = k \cdot \frac{1}{k} = 1$$

Let us prove a theorem expressing a sufficient condition for the passage to the limit under the integration sign to be legitimate.

Theorem VII.3.1 (H. Lebesgue). Let f_k ($k = 1, 2, \dots$) be bounded and measurable functions on a set $E \subset X$ (with $\mu E < +\infty$) and let there exist a constant $M > 0$ such that $|f_k(x)| \leq M$ for all k and for almost all $x \in E$. If $f_k \Rightarrow f$

on E^* and the function f is bounded on the set E relation (11) holds.

Proof. Replacing, if necessary, the functions f_k by equivalent functions we can at the very beginning suppose that $|f_k(x)| \leq M$ for all $x \in E$ ($k = 1, 2, \dots$). As we know, such a replacement does not affect the values of the integrals $\int_E f_k d\mu$ (see Theorem VII.2.5) and of course does not

violate the convergence in measure of f_k to f . Furthermore, passing, if necessary, to a greater value of M we can suppose that $|f|$ is bounded by the same constant M : $|f(x)| \leq M$ throughout E .**

We shall assume that $\mu E > 0$ since in the case $\mu E = 0$ relation (11) is trivial.

Let us take an arbitrary $\varepsilon > 0$, put $\eta = \frac{\varepsilon}{2\mu E}$ and construct the sets

$$E_k = E [|f_k(x) - f(x)| \geq \eta] \quad (k = 1, 2, \dots)$$

The definition of convergence in measure indicates that $\mu E_k \rightarrow 0$. Consequently

$$\mu E_k < \frac{\varepsilon}{4M} \quad \text{for } k \geq K$$

where K is sufficiently large. Taking $k \geq K$ we shall estimate the difference

$$\int_E f_k d\mu - \int_E f d\mu$$

Using the properties of the integral already established and, in particular, Theorem VII.2.4, we can write

$$\begin{aligned} \left| \int_E f_k d\mu - \int_E f d\mu \right| &\leq \int_E |f_k - f| d\mu = \\ &= \int_{E_k} |f_k - f| d\mu + \int_{E \setminus E_k} |f_k - f| d\mu \end{aligned} \quad (12)$$

* And, in particular, if $f_k(x) \rightarrow f(x)$ almost everywhere (see Theorem VI.5.2).

** By the way, Riesz' Theorem VI.5.3 shows directly that if $|f_k(x)| \leq M$ on E for all k then $|f(x)| \leq M$ almost everywhere on E .

Since $|f_k(x) - f(x)| \leq 2M$ throughout E we have

$$\int_{E_k} |f_k - f| d\mu \leq 2M \cdot \mu E_k < \frac{\varepsilon}{2}$$

On the other hand, $|f_k(x) - f(x)| < \eta$ on $E \setminus E_k$, and therefore

$$\int_{E \setminus E_k} |f_k - f| d\mu \leq \eta \cdot \mu(E \setminus E_k) \leq \eta \cdot \mu E = \frac{\varepsilon}{2}$$

Comparing the last inequality with (12) we readily see that

$$\left| \int_E f_k d\mu - \int_E f d\mu \right| < \varepsilon$$

which, by the arbitrariness of ε , proves relation (11).

In the next chapter we shall prove a stronger variant of Lebesgue's theorem on the passage to the limit under the integral sign.

§ 4. The Space S of Measurable Functions

Using the Lebesgue integral of bounded functions we can introduce a metric in the set of *all* measurable (and finite almost everywhere) functions defined on a set E with finite measure.

Let $E \subset X$ and $\mu E < +\infty$. By S we shall mean the set of all measurable functions f defined on E which are finite almost everywhere. We shall agree that equivalent functions are identified with each other, that is they are regarded as a single element of the set S . Thus, every element f of S is represented not by one definite function but by the corresponding class of all mutually equivalent functions. In particular, every element $f \in S$ can be represented by a function assuming finite values everywhere on E and thus we can think of S as consisting of only such functions.

Given two arbitrary elements $f, g \in S$, we define the distance between them by the formula

$$\rho(f, g) = \int_E \frac{|f - g|}{1 + |f - g|} d\mu \quad (13)$$

The integrand being bounded (its values lie between 0 and 1) and measurable, the integral always exists. If f or g (or both) is replaced by an equivalent function belonging to S the distance $\rho(f, g)$ determined by formula (13) does not change, that is $\rho(f, g)$ is independent of the particular choice of the functions representing the elements f and g belonging to the corresponding classes of all the possible equivalent functions.

Let us show that the function $\rho(f, g)$ determined by formula (13) satisfies all the axioms of distance in a metric space. It is clear that $\rho(f, f) = 0$. From Theorem VII.2.6 it follows that if $\rho(f, g) = 0$ the integrand in (13) is a function equivalent to 0, that is $f \sim g$; this exactly means that f and g as elements of the set S are equal (i.e. they coincide). The symmetry axiom obviously holds for $\rho(f, g)$.

To establish the triangle inequality we first derive an auxiliary relation. Let $0 \leq \alpha \leq \beta$. Then $\alpha + \alpha\beta \leq \beta + \alpha\beta$. Dividing both sides by $(1 + \alpha)(1 + \beta)$ we obtain

$$\frac{\alpha}{1 + \alpha} \leq \frac{\beta}{1 + \beta} \quad (14)$$

Now, for any two real numbers a and b we obtain, with the aid of (14), the inequalities

$$\begin{aligned} \frac{|a+b|}{1+|a+b|} &\leq \frac{|a|+|b|}{1+|a|+|b|} = \\ &= \frac{|a|}{1+|a|+|b|} + \frac{|b|}{1+|a|+|b|} \leq \frac{|a|}{1+|a|} + \frac{|b|}{1+|b|} \end{aligned}$$

Taking three arbitrary functions f , g and h belonging to S and putting

$$a = f(x) - h(x) \text{ and } b = h(x) - g(x)$$

we derive

$$\frac{|f(x) - g(x)|}{1+|f(x) - g(x)|} \leq \frac{|f(x) - h(x)|}{1+|f(x) - h(x)|} + \frac{|h(x) - g(x)|}{1+|h(x) - g(x)|}$$

On integrating termwise the last relation over E we finally obtain

$$\rho(f, g) \leq \rho(f, h) + \rho(h, g)$$

Thus, the set S (equipped with metric (13)) is a metric space.

It is readily seen from formula (13) that for any two elements f and $g \in S$ the distance $\rho(f, g)$ is equal to the distance from the element $f - g$ to the element θ :

$$\rho(f, g) = \rho(f - g, \theta)$$

(θ is the zero element of the space S represented not only by the function identically equal to zero but also to every function equivalent to 0).

Theorem VII.4.1. *The convergence with respect to the distance in S coincides with the convergence in measure.*

Proof. Let $f_k \Rightarrow f$ ($f, f_k \in S$). Then, given any $\varepsilon > 0$, we have

$$\mu E \left[\frac{|f_k(x) - f(x)|}{1 + |f_k(x) - f(x)|} \geq \varepsilon \right] \leq \mu E [|f_k(x) - f(x)| \geq \varepsilon]$$

Therefore the sequence $\left\{ \frac{|f_k - f|}{1 + |f_k - f|} \right\}$ converges in measure to 0. Consequently, by Theorem VII.3.1,

$$\rho(f_k, f) = \int_E \frac{|f_k - f|}{1 + |f_k - f|} d\mu \rightarrow 0$$

On the other hand, if $\{f_k\}$ is not convergent in measure to f there exists some $\varepsilon > 0$ such that

$$\mu E [|f_k(x) - f(x)| \geq \varepsilon] \not\rightarrow 0$$

This means that there is $\delta > 0$ such that the inequality

$$\mu E [|f_{k_i}(x) - f(x)| \geq \varepsilon] \geq \delta$$

holds for an infinite number of the values k_i of the index k . According to inequality (14), if $|f_{k_i}(x) - f(x)| \geq \varepsilon$ then

$$\frac{|f_{k_i}(x) - f(x)|}{1 + |f_{k_i}(x) - f(x)|} \geq \frac{\varepsilon}{1 + \varepsilon}$$

and therefore

$$\rho(f_{k_i}, f) = \int_E \frac{|f_{k_i} - f|}{1 + |f_{k_i} - f|} d\mu \geq \frac{\delta \varepsilon}{1 + \varepsilon}$$

Consequently, $\rho(f_{k_i}, f) \not\rightarrow 0$.

Theorem VII.4.2. *S is a complete metric space.*

Proof. Let $\{f_k\}$ be a Cauchy (fundamental) sequence of elements of S , that is $\rho(f_k, f_l) \rightarrow 0$ for $k, l \rightarrow \infty$. Using the argument analogous to the one used above we readily conclude that

$$\mu E[|f_k(x) - f_l(x)| \geq \varepsilon] \xrightarrow[k, l \rightarrow \infty]{} 0$$

for any $\varepsilon > 0$. By Lemma VI.5.1, there is a subsequence $\{f_{k_i}\}$ of $\{f_k\}$ convergent almost everywhere to a measurable limit function f with finite values. By Theorem VI.5.2, convergence almost everywhere implies convergence in measure; therefore $f_{k_i} \rightarrow f$ with respect to the distance in S . By Lemma III.4.1, it follows that $f_k \rightarrow f$. Thus, every Cauchy sequence in S has a limit, which means that the space S is complete.

Let us consider the special case when $X = R_n$ and μ is the Lebesgue measure (as before, we suppose that $\mu E < +\infty$). In this case *the space S is separable*. To prove this property it suffices to show that the set $\mathcal{P}$ of all algebraic polynomials in n variables with rational coefficients is everywhere dense in S (see the end of Sec. I.4).

Let $f \in S$. By Theorem VI.6.5, there exists a sequence of functions continuous on R_n which is convergent to f almost everywhere on E and thus is convergent with respect to the metric in S . Consequently, given an arbitrary $\varepsilon > 0$, there is a function g continuous throughout R_n whose restriction to the set E (we retain the same notation g for the restriction) satisfies the inequality $\rho(f, g) < \frac{\varepsilon}{2}$. Furthermore, Weierstrass' theorem readily shows that for any continuous function φ on R_n there exists a sequence of polynomials belonging to $\mathcal{P}$ which is convergent to $\varphi(x)$ for all $x \in R_n$.^{*} Therefore, by analogy with the above, we conclude that for the function g there is a polynomial Q belonging to $\mathcal{P}$ such that $\rho(g, Q) < \frac{\varepsilon}{2}$ with respect to the metric of the space S on the set E .

^{*} To obtain such a sequence we should choose, for every natural number p , a polynomial Q_p belonging to $\mathcal{P}$ such that $|Q_p(x) - \varphi(x)| < \frac{1}{p}$ for all x 's belonging to the parallelepiped $[-p, p; -p, p; \dots, -p, p]$.

Consequently, $\rho(f, Q) < \varepsilon$, which proves the separability of S .

Now we proceed to consider a rather general notion introduced by the Soviet mathematicians P. S. Alexandrov (born 1896) and P. S. Uryson (1898-1924).

Definition. Let Y be a set for which a certain kind of convergence is defined, that is let there be an indication which of the sequences of elements of Y are regarded as convergent and what their limits are. Let this type of convergence be called (l) -convergence (written $y_k \xrightarrow{(l)} y$).^{*} A sequence $\{y_k\}$ of elements of Y is said to be $(*)$ -convergent to $y \in Y$ (written $y_k \xrightarrow{(*)} y$) if its every subsequence $\{y_{k_i}\}$ contains a subsequence $\{y_{k_{i_j}}\}$ such that $y_{k_{i_j}} \xrightarrow{l} y$. More precisely, we should speak of $(*)$ -convergence relative to the given type of (l) -convergence.

If Y is a metric space the $(*)$ -convergence relative to the convergence with respect to the distance coincides with the convergence with respect to the distance. Indeed, if $y_k \xrightarrow{l} y$ with respect to the distance we have $y_{k_i} \xrightarrow{l} y$ for any subsequence $\{y_{k_i}\}$ and then $y_k \xrightarrow{(*)} y$. If $y_k \not\xrightarrow{l} y$ with respect to the distance there is some $\varepsilon > 0$ such that $\rho(y_{k_i}, y) \geq \varepsilon$ for an infinite number of values $k_1 < k_2 < \dots < k_i < \dots$ of the index k . Then there can be no subsequence of the subsequence $\{y_{k_i}\}$ convergent to y with respect to the distance.

Thus, if a notion of (l) -convergence is defined in a set Y then, for the set Y to be metrizable so that the convergence with respect to the distance coincides with the (l) -convergence it is necessary (but not sufficient) that the (l) -convergence should coincide with the corresponding type of $(*)$ -convergence.

From Lebesgue's Theorem VI.5.2 and Riesz' Theorem VI.5.3 it readily follows that if Y is the set S of measurable functions defined on a set $E \subset X$ with $\mu E < +\infty$ and if the (l) -convergence in S is understood as the convergence

^{*} For example, Y can be a metric space in which the notion of (l) -convergence is defined as convergence with respect to the distance. Also, Y can be the set S of measurable functions and (l) -convergence in it can be defined as convergence almost everywhere, and the like.

almost everywhere the corresponding type of $(*)$ -convergence is the convergence in measure which may not coincide with the convergence almost everywhere (see the example in Sec. VI.5). Comparing this with the foregoing result we conclude that in this case *it is impossible* to equip S with a metric so that the convergence with respect to the distance coincides with the convergence almost everywhere. Thus, convergence in measure is in a certain sense simpler than convergence almost everywhere since the former can be realized as convergence with respect to a metric.

In connection with our study of the space S let us dwell on a general notion related to arbitrary sets. Suppose that for some of the elements $x, y \in M$ of a set M there is defined the notion of *equivalence* (written $x \sim y$). Let the following conditions be fulfilled:

- (i) if $x \sim y$ then $y \sim x$ (the *symmetry*);
- (ii) $x \sim x$ for any $x \in M$ (the *reflexivity*);
- (iii) if $x \sim y$ and $y \sim z$ then $x \sim z$ (the *transitivity*).

In this case we say that there is an *equivalence relation* in the set M . It is evident that the equivalence relation which was introduced for functions belonging to the space S possesses all the enumerated properties.

We shall show that *if an equivalence relation is introduced in a set M this set can be represented as the union of a family of pairwise disjoint subsets of M such that the elements of each of these subsets are mutually equivalent while the elements entering into different subsets belonging to this family are not equivalent to each other*. These subsets are called *equivalence classes* and the above representation is called a *partition* or *decomposition of M into equivalence classes*.

To prove the possibility of this representation let us denote by K_x the collection of all elements of M equivalent to x where x is an arbitrary element of M . It is clear that $K_x \neq \emptyset$ since $x \in K_x$. Any two elements $y, z \in K_x$ are mutually equivalent. Indeed, we have $y \sim x$ and $z \sim x$ whence, by the symmetry and the transitivity of the equivalence relation, we immediately obtain $y \sim z$. Furthermore, if $y \in K_x$ and u is an element equivalent to y the transitivity implies that $u \sim x$, i.e. $u \in K_x$; if $x \sim y$ then also $K_x = K_y$. It now remains to show that if any two classes K_x and K_y do not coincide they are disjoint. To

this end, let us suppose that there exists $z \in K_x \cap K_y$. Then $z \sim x$ and $z \sim y$, and consequently $x \sim y$ and $K_x = K_y$, which contradicts the hypothesis. Thus, the family of all *different* subsets K_x gives the desired decomposition into equivalence classes.

Now suppose that all the mutually equivalent elements of the set M are (mutually) "identified". Strictly speaking, this means that instead of M we consider the new set M' (called a *factor set* or a *quotient set* of M) whose elements are the (equivalence) classes of mutually equivalent elements. When there is no danger of misinterpretation we sometimes prefer, as was done in the case of the space S , to retain for the set M' with identified elements the same notation by the letter M and to think of its elements not as classes but as certain representing elements of the original set. But then each of the elements of the (new) set M interpreted in this new way is represented not only by a certain "individual" element of the original set but also by any other element equivalent to the former (i.e. by any element belonging to the corresponding class).

The space S treated in this section is interpreted in exactly this way. Speaking more strictly we should say that here by S is in fact meant not the set of all measurable functions finite almost everywhere on E but the factor set of that set.

Finally, suppose that there are some operations on elements defined for the set M with respect to which M is closed, that is these operations on elements of M result in "objects" which again are elements of M . For instance, an operation of this kind can be addition of elements. If these operations are "coherent" with the equivalence relation in the sense that when any elements are replaced by equivalent elements the results of the operations on them also go into equivalent elements the operations on elements of M can be in a natural manner extended to the factor set and can be regarded as operations on equivalence classes. It is advisable to recall what has been said here in the study of the space L in Sec. VIII.5.

Let us come back to the space S .

Theorem VII.4.3. *For any countable set of nonnegative functions $f_k \in S$ ($k = 1, 2, \dots$) there exist numbers $\alpha_k > 0$*

and a function $h \in \mathcal{S}$ such that $\alpha_k f_k(x) \leq h(x)$ almost everywhere on E for any k .

Proof. For every k and any $x \in E$ we have

$$\lambda f_k(x) \xrightarrow{\lambda \rightarrow +0} 0$$

(we remind the reader that the functions f_k can be assumed to be finite everywhere). Consequently, $\rho(\lambda f_k, \theta) \rightarrow 0$ for $\lambda \rightarrow +0$, and therefore there are $\alpha_k > 0$ ($k = 1, 2, \dots$) such that $\rho(\alpha_k f_k, \theta) < \frac{1}{2^k}$. Let us put

$$h_k = \sum_{l=1}^k \alpha_l f_l \quad (k = 1, 2, \dots)$$

Then

$$\rho(h_{k+1}, h_k) = \rho(h_{k+1} - h_k, \theta) = \rho(\alpha_{k+1} f_{k+1}, \theta) < \frac{1}{2^{k+1}}$$

and

$$\rho(h_m, h_k) \leq \sum_{l=k}^{m-1} \rho(h_{l+1}, h_l) < \sum_{l=k}^{m-1} \frac{1}{2^{l+1}} < \frac{1}{2^k}$$

for any $m > k$. Thus, $\{h_k\}$ is a Cauchy sequence, and, by the completeness of the space $\mathcal{S}$, there exists $h \in \mathcal{S}$ such that $\rho(h_k, h) \rightarrow 0$. This means that $h_k \Rightarrow h$. Therefore, by Riesz' Theorem VI.5.3, there is a subsequence $\{h_{k_i}\}$ such that $h_{k_i}(x) \rightarrow h(x)$ almost everywhere. The functions h_k forming an increasing sequence, $\lim h_k(x)$ exists at every point. Together with the above the last conclusion shows that $h(x) = \lim h_k(x) = \sup h_k(x)$ almost everywhere. Consequently, for any k we have

$$\alpha_k f_k(x) \leq h_k(x) \leq h(x)$$

almost everywhere on E .

Now we can strengthen Theorem VI.6.2 on a convergence regulator proved in the previous chapter.

Theorem VII.4.4 (On the Existence of a Common Convergence Regulator). *If there are functions*

$$f_k^{(p)} \in \mathcal{S} \quad (k, p = 1, 2, \dots)$$

such that

$$f_k^{(p)}(x) \xrightarrow{k \rightarrow \infty} 0$$

for each p almost everywhere on E there exist constants $\beta_{p,k}$ and a measurable, finite almost everywhere and nonnegative function $g(x)$ (a convergence regulator common for all the sequences $\{f_k^{(p)}\}$) such that $\beta_{p,k} > 0$, $\beta_{p,k} \rightarrow 0$ for $k \rightarrow \infty$ and

$$|f_k^{(p)}(x)| \leq \beta_{p,k} g(x) \quad (p, k = 1, 2, \dots)$$

Proof. Let $g_p \in S$ ($p = 1, 2, \dots$) be the function (of the type of g) indicated in Theorem VI.6.2 for the sequence of functions $f_k^{(p)}$ ($k = 1, 2, \dots$). By the foregoing theorem, it is possible to choose a new function $g \in S$ such that

$$\alpha_p g_p(x) \leq g(x) \quad \text{for almost all } x \in E$$

for some $\alpha_p > 0$. Then g can serve as a common convergence regulator whose existence has thus been proved.

Indeed, for each p there is a sequence $\lambda_k^{(p)} \xrightarrow[k \rightarrow \infty]{} 0$ such that

$$|f_k^{(p)}(x)| \leq \lambda_k^{(p)} g_p(x) \quad \text{almost everywhere on } E$$

Now it follows that

$$|f_k^{(p)}(x)| \leq \frac{\lambda_k^{(p)}}{\alpha_p} g(x)$$

almost everywhere on E , and it only remains to put $\beta_{p,k} = \frac{\lambda_k^{(p)}}{\alpha_p}$.

The theorem below was suggested by R. M. Fréchet.

Theorem VII.4.5 (On the Diagonal Sequence). Let $f_k^{(p)}$, f_p ($k, p = 1, 2, \dots$) and f be functions belonging to S such that

(i) $f_k^{(p)}(x) \xrightarrow[k \rightarrow \infty]{} f_p(x)$ for every $p = 1, 2, \dots$ almost everywhere on E ;

(ii) $f_p(x) \xrightarrow[p \rightarrow \infty]{} f(x)$ almost everywhere on E .

Then there exists an increasing sequence k_p of values of the index k such that

$$f_{k_p}^{(p)}(x) \xrightarrow[p \rightarrow \infty]{} f(x) \quad \text{almost everywhere on } E.$$

If all the functions $f_k^{(p)}$ are arranged as the infinite matrix

$$\begin{array}{ccccccc} f_1^{(1)}, & f_2^{(1)}, & \dots, & f_k^{(1)}, & \dots & & \\ f_1^{(2)}, & f_2^{(2)}, & \dots, & f_k^{(2)}, & \dots & & \\ \dots & \dots & \dots & \dots & \dots & \dots & \\ f_1^{(p)}, & f_2^{(p)}, & \dots, & f_k^{(p)}, & \dots & & \\ \dots & \dots & \dots & \dots & \dots & \dots & \end{array}$$

the required sequence $\{f_{k_p}^{(p)}\}$ is obtained as follows: from the first row we take the element with a number k_1 , from the second row the element with a number $k_2 > k_1$ whose column is to the right of that of $f_{k_1}^{(1)}$ and so on. A sequence chosen in this fashion is called *diagonal*.

Proof. Let us apply the foregoing theorem to all the sequences of the form $\{f_k^{(p)} - f_p\}$ and $\{f_p - f\}$ and find the corresponding function of the type of g indicated in the theorem. Next we find $\alpha_p \rightarrow 0$ such that, for any p ,

$$|f_p(x) - f(x)| \leq \alpha_p g(x) \text{ almost everywhere on } E$$

Further, for every p we choose a value k_p of the index k such that

$$|f_{k_p}^{(p)}(x) - f_p(x)| \leq \alpha_p g(x) \text{ almost everywhere on } E$$

In this construction k_p 's can be chosen so that $k_1 < k_2 < \dots < k_p < \dots$. Now, from the inequality

$$|f_{k_p}^{(p)}(x) - f(x)| \leq 2\alpha_p g(x)$$

which holds almost everywhere on E , it follows that $f_{k_p}^{(p)}(x) \rightarrow f(x)$ also almost everywhere on E .

Theorem VII.4.5 enables us to draw the following conclusion: if we tried to construct in a Euclidean space (or on its measurable subset) an analogue of Baire's classification of functions proceeding not from pointwise convergence but from convergence almost everywhere this would only result in the functions of the first class. Indeed, any function representable as a limit almost everywhere of a sequence of functions of the first class would itself belong to the first class, and thus there would not appear higher classes.

The metric of the space S enables us to prove an interesting property of S which is by far not obvious. Let us introduce a *partial ordering* in the space S by putting $f \leq g$ ($f, g \in S$) if $f(x) \leq g(x)$ almost everywhere on E .^{*} If $T \subset S$ is a subset of S and there exists a function $g \in S$ such that $f \leq g$ for any $f \in T$ ^{**} the set T is said to be *bounded above*, g being called its *upper bound*. It turns out that the classical theorem on the existence of the least upper bound (supremum) for a set of numbers bounded above can be generalized to the space S . Below is the exact statement of this generalization.

Theorem VII.4.6. *Let T be an arbitrary nonempty set of functions, belonging to the space S , bounded above. Then there exists a function $h \in S$ such that*

- (i) h is an upper bound of the set T ;
 - (ii) if g is any other upper bound of the set T then $h \leq g$.
- (Such a function h is called the *least upper bound* or the *supremum* of the set T .)

It should be noted that if the set T is countable the desired function h can be defined by the formula

$$h(x) = \sup_{f \in T} f(x) \quad *** \quad (15)$$

But if the set T is uncountable then, generally speaking, such a definition of h is impossible since in this case the function determined by formula (15) can even be nonmeasurable. For instance, if $E = [0, 1] \subset \mathbb{R}_1$, μ is the Lebesgue measure, $H \subset E$ is a nonmeasurable subset and T consists of the characteristic functions of all the single-element subsets of the set H then $\sup_{f \in T} f(x)$ is the characteristic function of the set H and therefore it does not belong to S . At the same time, it is easy to understand that in this example the

^{*} In such a case we speak of *partial ordering* and say that S is *partially ordered* since the ordering we have introduced only makes it possible to compare some of the functions but not any two functions belonging to S .

^{**} In other words, for any $f \in T$ the inequality $f(x) \leq g(x)$ holds almost everywhere on E ; the set of measure zero on which this inequality can be violated may depend on f .

^{***} To construct the function h so that it is finite everywhere we should add to formula (15) the condition that at the points x of the set (of measure zero) on which $\sup f(x) = +\infty$ the values of $h(x)$ are put equal to a finite constant.

supremum of the set T is the function $h(x) \equiv 0$ (or any function equivalent to it).

Proof of the theorem. Let us choose an arbitrary function $f_0 \in T$ and consider the set T' consisting of all the functions of the form $\varphi = \max(f, f_0)$, that is $\varphi(x) = \max[f(x), f_0(x)]$ ($f \in T$).^{*} It is clear that the sets T and T' have the same upper bounds, and therefore it is sufficient to prove the existence of the supremum for the set T' . Let us consider the set T'' of all functions of the form $\varphi - f_0$ ($\varphi \in T'$). This set consists of nonnegative functions, and its upper bounds are the functions of the form $g - f_0$ where g 's are the upper bounds of the set T' . If we show that T'' possesses the supremum (denote it by ω) the function $h = \omega + f_0$ will be the supremum of T' and thus of T as well. Hence, at the very beginning, we can assume, without loss of generality, that T consists of nonnegative functions. Furthermore, let us also assume that for any finite subset of functions $f_1, f_2, \dots, f_p$ belonging to T the function $f = \max(f_1, f_2, \dots, f_p) \in T$. This is allowable since the additional inclusion in T of all functions of this type does not affect its upper bounds, and hence the latter assumption does not violate generality either.

Taking an arbitrary function $f \in S$ nonnegative almost everywhere we put

$$m(f) = \rho(f, \theta) = \int_E \frac{f}{1+f} d\mu$$

Inequality (14) implies that if $0 \leq f_1(x) \leq f_2(x)$ almost everywhere on E ($f_1, f_2 \in S$) then $m(f_1) \leq m(f_2)$. In particular, if g is an upper bound of the set T then $m(f) \leq m(g)$ for any $f \in T$. Let us put

$$M = \sup_{f \in T} m(f) \quad (M < +\infty)$$

and find a functional sequence $f_k \in T$ such that $m(f_k) \rightarrow M$. Replacing, if necessary, each function f_k by $\max(f_1, f_2, \dots, f_k)$ we can also assume at the very beginning that f_k 's form an increasing sequence.

^{*} The measurability of the function φ was proved in Sec. VI.3.

Now let $h(x) = \sup f_k(x) = \lim f_k(x)$.^{*} Then, by Theorem VII.3.1,

$$m(f_k) = \int_E \frac{f_k}{1+f_k} d\mu \rightarrow \int_E \frac{h}{1+h} d\mu = m(h)$$

and therefore $m(h) = M$. Moreover, $h(x) \leq g(x)$ almost everywhere on E for any upper bound g of the set T , that is h satisfies condition (ii). It remains to check condition (i).

Let us take an arbitrary function $f^* \in T$ and put $f_k^* = \max(f_k, f^*)$ and $h^* = \max(h, f^*)$. Then $f_k^* \in T$ and $\{f_k^*(x)\}$ is an increasing sequence convergent to $h^*(x)$. Therefore $m(f_k^*) \rightarrow m(h^*)$ and, consequently, $m(h^*) \leq M$. On the other hand, $h^* \geq h$, and thus $m(h^*) \geq m(h) = M$ whence $m(h^*) = M$. Hence,

$$\int_E \left(\frac{h^*}{1+h^*} - \frac{h}{1+h} \right) d\mu = m(h^*) - m(h) = 0$$

Since the integrand in this integral is nonnegative, Theorem VII.2.6 indicates that this function is equal to zero almost everywhere, and hence $h^* \sim h$. But we have $f^*(x) \leq h^*(x)$ for all $x \in E$; consequently $f^*(x) \leq h(x)$ almost everywhere on E . Thus, $f^* \leq h$ for any $f^* \in T$, that is h is the supremum of the set T .

^{*} Also cf. the footnote on p. 177 concerning formula (15).

Summable Functions

§ 1. The Definition of a Summable Function*

In this section we shall generalize the concept of the Lebesgue integral to the case when the integrand can be an unbounded function and the domain of integration can have an infinite measure. As before, we consider an abstract space X with a measure μ defined on a σ -algebra of $\mathcal{S}$ of measurable sets, the measure μ being σ -finite. We shall begin with the definition of the integral of a nonnegative function.

Definition. Let a function $f(x) \geq 0$ be measurable and finite almost everywhere on a set $E \subset X$. Taking all the possible measurable subsets $e \subset E$ having a finite measure on which the function f is bounded** we put

$$\int_E f d\mu = \sup_e \int_e f d\mu \quad (1)$$

We thus have $\int_E f d\mu \geq 0$. Then, if $\int_E f d\mu < +\infty$, the function f is said to be *summable* with respect to the measure μ or, simply, μ -*summable* on the set E .

It is obvious that if a nonnegative function f is bounded and $\mu E < +\infty$ then, among all the integrals entering into the right-hand side of formula (1), there exists the greatest

* See the footnote on p. 152.—Tr.

** For such sets e the integral $\int_e f d\mu$ was already defined in Sec. VII.1. Sets $e \subset E$ with finite measure on which f is bounded always exist; for instance, such is the empty set. But it can be shown that the σ -finiteness of the measure μ implies that, except the case when $\mu E = 0$ and $f(x) = +\infty$ everywhere on E (in this case we can also say that f is almost everywhere finite), there also exist nonempty sets $e \subset E$ with finite measure on which f is bounded.

one which is the integral over the set E .^{*} Consequently, in this case the integral of f defined by formula (1) coincides with the integral defined in Chapter VII with the aid of the Lebesgue-Darboux sums. Hence, *on a set with a finite measure, every bounded nonnegative measurable function is summable*. A bounded function on a set with infinite measure may not be summable. For instance, if $f(x) \equiv c > 0$ where c is a constant and $\mu E = +\infty$ then $\int_E f d\mu = +\infty$.^{**}

It follows immediately from the definition that if $E' \subset E$ where E' is a measurable set then

$$\int_{E'} f d\mu \leq \int_E f d\mu$$

Consequently, if f is summable on E it is summable on E' as well.

If $\mu E = 0$ every nonnegative measurable function f is summable on E and $\int_E f d\mu = 0$. Indeed, in this case all the integrals entering into the right-hand side of relation (1) are equal to zero since $\mu e = 0$, and therefore the left-hand side also turns into zero.

If $f(x) \equiv 0$ on E and E is an arbitrary measurable set then, obviously, $\int_E f d\mu = 0$.

Let us extend the definition of the concept of the integral given above to the case of a function $f(x) \geq 0$ measurable on a set $E \subset X$ for which $\mu E [f(x) = +\infty] > 0$. In this case we put, by definition, $\int_E f d\mu = +\infty$.

If $f(x)$ is a measurable function on a set $E \subset X$ and $f(x) \leq 0$ we put

$$\int_E f d\mu = - \int_E |f| d\mu$$

^{*} See Corollary 3 of the theorem in Sec. VII.2.2.

^{**} Indeed, it follows from the σ -finiteness of the measure μ that the set E contains subsets e with an arbitrarily large finite measure and therefore

$$\int_E f d\mu = \sup_{e \subset E} \int_e f d\mu = \sup_e c \cdot \mu e = +\infty$$

where the integral on the right-hand side has already been defined. In this case we say that the function f is *summable* if the integral $\int_E |f| d\mu$ assumes a finite value, that is if the function $|f|$ is summable.

For a function f which takes on both positive and negative values the notion of the integral is introduced as follows. Let f be measurable on a set E . Putting

$$E_1 = E [f(x) \geq 0] \quad \text{and} \quad E_2 = E [f(x) < 0]$$

we thus break E into two measurable sets. Now the integral of f over the set E is defined by the formula

$$\int_E f d\mu = \int_{E_1} f d\mu + \int_{E_2} f d\mu = \int_{E_1} |f| d\mu - \int_{E_2} |f| d\mu \quad (2)$$

It should be taken into account that formula (2) is meaningful only if at least one of the integrals $\int_{E_1} f d\mu$ and $\int_{E_2} f d\mu$ is finite. If both integrals are infinite, that is

$$\int_{E_1} f d\mu = +\infty, \quad \int_{E_2} f d\mu = -\infty$$

the integral $\int_E f d\mu$ makes no sense.

The function f is called *summable* on the set E if f (or, which is the same, $|f|$) is summable on each of the sets E_1 and E_2 , that is if $\int_E |f| d\mu$ has a finite value.

Apparently, if a function is summable on a set E it is also summable on its every measurable subset. If $\mu E = 0$, every measurable function f defined on E is summable and

$$\int_E f d\mu = 0$$

The last two assertions are implied by the fact that they hold for nonnegative functions.

We again note that if a function f is bounded and measurable on a set E with finite measure then, since the integrals of this function taken over the sets E_1 and E_2 considered

in this chapter assume the same values as in Chapter VII, the value of the integral $\int_E f d\mu$ defined by formula (2) coincides with that of the integral defined in Chapter VII. Thus, *if f is a bounded measurable function on a set E with $\mu E < +\infty$, it is summable on this set.*

§ 2. Lemmas on Integrals of Nonnegative Functions

Here we shall prove some lemmas used in the following sections. These lemmas generalize the corresponding assertions proved in Sec. VII.2 for integrals of bounded functions over sets of finite measure. All the functions below are supposed to be finite almost everywhere, and we do not stipulate this condition in the statements of the lemmas.

Lemma VIII.2.1 (On the σ -additivity of the Integral). *Let E be a set representable as a finite or countable union of disjoint measurable sets E_j : $E = \bigcup_j E_j$. Then for any non-negative function f measurable on the set E we have*

$$\int_E f d\mu = \sum_j \int_{E_j} f d\mu \quad (3)$$

In particular, *if the number of the sets E_j is finite and f is summable on each E_j , the function f is summable on E .*

Proof. If at least one of the integrals $\int_{E_j} f d\mu$ is equal to $+\infty$ then $\int_E f d\mu = +\infty$, and equality (3) obviously holds.

Suppose that all the integrals $\int_{E_j} f d\mu < +\infty$. Let us take an arbitrary measurable set $e \subset E$ with finite measure on which f is bounded and put $e_j = e \cap E_j$. Then the σ -additivity of the integral of a bounded function implies

$$\int_e f d\mu = \sum_j \int_{e_j} f d\mu \leq \sum_j \int_{E_j} f d\mu$$

Taking the supremum of the left-hand side we obtain

$$\int_E f d\mu \leq \sum_j \int_{E_j} f d\mu$$

Now let us take the first k sets of E_j and an arbitrary $\varepsilon > 0$ and choose, for each $j = 1, 2, \dots, k$, a measurable set $e_j \subset E_j$ with finite measure, on which f is bounded, such that

$$\int_{e_j} f d\mu > \int_{E_j} f d\mu - \frac{\varepsilon}{k}$$

Putting $e = \bigcup_{j=1}^k e_j$ we see that $e \subset E$, $\mu e < +\infty$ and f is bounded on e . Then, obviously,

$$\int_e f d\mu > \sum_{j=1}^k \int_{E_j} f d\mu - \varepsilon$$

that is

$$\sum_{j=1}^k \int_{E_j} f d\mu < \int_e f d\mu + \varepsilon \leq \int_E f d\mu + \varepsilon$$

Since ε is arbitrary it follows that

$$\sum_{j=1}^k \int_{E_j} f d\mu \leq \int_E f d\mu$$

Finally, making k tend to infinity if E is broken up into an infinite number of sets E_j or putting k equal to the number of these sets if their collection is finite we arrive at the inequality

$$\sum_j \int_{E_j} f d\mu \leq \int_E f d\mu$$

which together with the opposite inequality established above implies equality (3).

Remark. The above lemma implies a corollary analogous to Corollary 2 of Theorem VII.2.2.

From Lemma VIII.2.1 it readily follows that formula (2) in the definition of the integral of a function f assuming

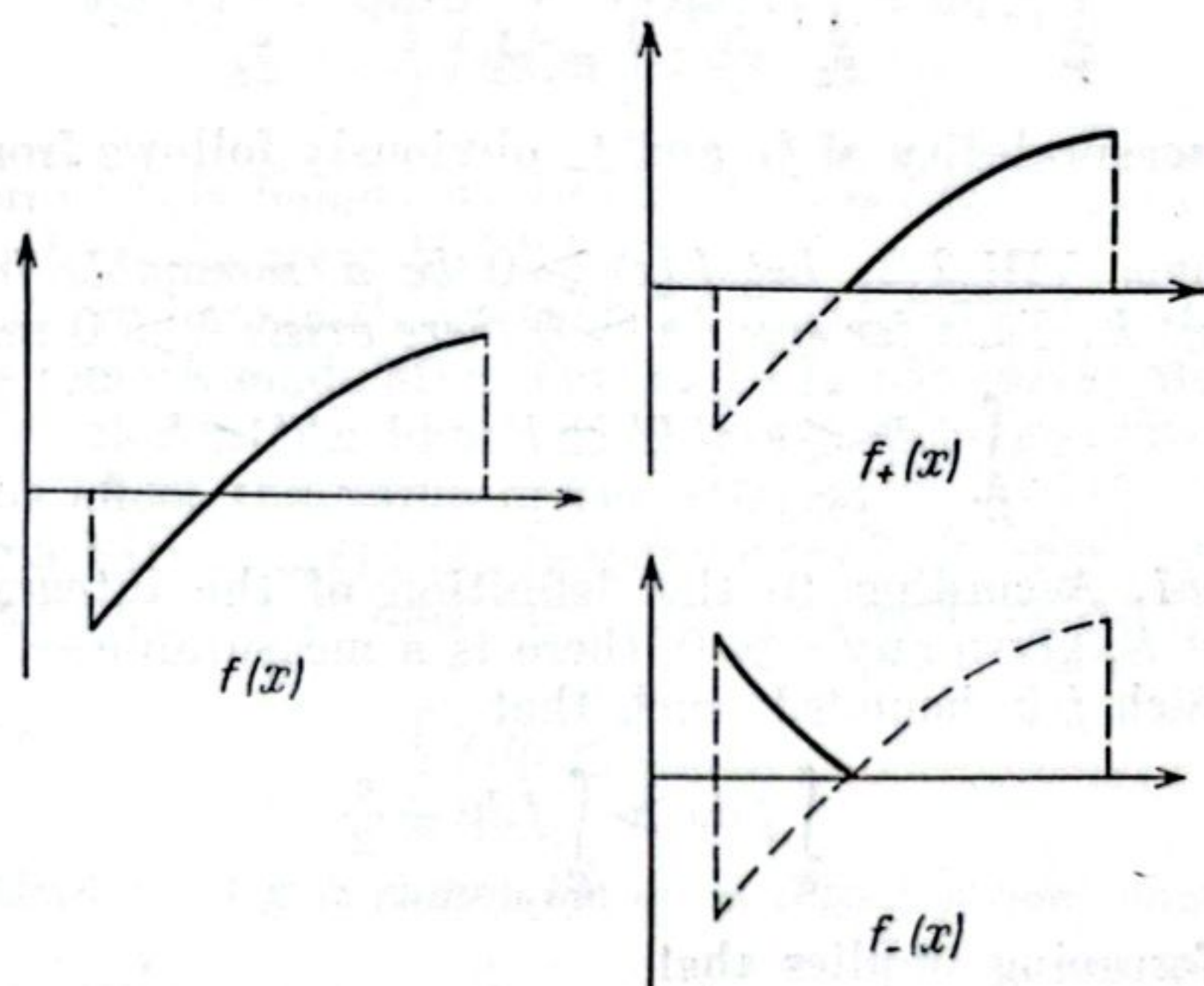


Fig. 18

values of different signs can be replaced by the equivalent formula

$$\int_E f d\mu = \int_E f_+ d\mu - \int_E f_- d\mu \quad (4)$$

where (see Fig. 18)

$$f_+(x) = \begin{cases} f(x) & \text{if } f(x) \geq 0 \\ 0 & \text{if } f(x) < 0 \end{cases}$$

and

$$f_-(x) = \begin{cases} |f(x)| & \text{if } f(x) \leq 0 \\ 0 & \text{if } f(x) > 0 \end{cases}$$

Indeed, if E_1 and E_2 are the sets meant in formula (2) then

$$\int_E f_+ d\mu = \int_{E_1} f d\mu + \int_{E \setminus E_1} 0 \cdot d\mu = \int_{E_1} f d\mu$$

and

$$\int_E f_- d\mu = \int_{E_2} |f| d\mu + \int_{E \setminus E_2} 0 \cdot d\mu = \int_{E_2} |f| d\mu$$

(the measurability of f_+ and f_- obviously follows from that of f).

Lemma VIII.2.2. *Let $f(x) \geq 0$ be a summable function on a set E . Then for every $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$\int_{E'} f d\mu < \varepsilon \text{ if } E' \subset E \text{ and } \mu E' < \delta$$

Proof. According to the definition of the integral over the set E , given any $\varepsilon > 0$, there is a measurable set $e \subset E$, on which f is bounded, such that

$$\int_e f d\mu > \int_E f d\mu - \frac{\varepsilon}{2}$$

The foregoing implies that

$$\int_E f d\mu = \int_e f d\mu + \int_{E \setminus e} f d\mu$$

Consequently

$$\int_{E \setminus e} f d\mu < \frac{\varepsilon}{2} \quad (5)$$

Let $0 \leq f(x) < M$ on the set e . Let us put $\delta = \frac{\varepsilon}{2M}$ and let $E' \subset E$ and $\mu E' < \delta$. Representing the set E' in the form $E' = E_1 \cup E_2$ where $E_1 = E' \cap e$ and $E_2 = E' \cap (E \setminus e)$ we can write, by Theorem VII.2.1,

$$\int_{E_1} f d\mu \leq M \cdot \mu E_1 \leq M \cdot \mu E' < \frac{\varepsilon}{2}$$

while inequality (5) implies

$$\int_{E_2} f d\mu \leq \int_{E \setminus e} f d\mu < \frac{\varepsilon}{2}$$

Therefore, $\int_{E'} f d\mu < \varepsilon$.

Lemma VIII.2.3. *If f and g are measurable on a set E , $f(x) \geq 0$, $g(x) \geq 0$ and $f \sim g$ then*

$$\int_E f d\mu = \int_E g d\mu$$

To prove this lemma we literally repeat the argument of the proof of Theorem VII.2.5.

The above lemma shows that, as in the foregoing chapter (see the remark made after Theorem VII.2.5), a variation of values of the integrand function on a set of measure zero does not affect the value of the integral.*

Lemma VIII.2.4. *If $f(x) \geq 0$ and $g(x) \geq 0$ are measurable functions on a set E and $f(x) \leq g(x)$ almost everywhere on E then*

$$\int_E f d\mu \leq \int_E g d\mu \quad (6)$$

In particular, if g is summable on E then f is also summable on E .

Proof. Without loss of generality we can suppose that $f(x) \leq g(x)$ for all $x \in E$.

Let us take an arbitrary set $e \subset E$ with $\mu e < +\infty$ on which f is bounded and consider the following two cases: (a) g is bounded on e , and (b) g is unbounded on e . In the former case, according to the properties of the integral of a bounded function and the definition of the integral of an unbounded function, we have

$$\int_e f d\mu \leq \int_e g d\mu \leq \int_E g d\mu$$

In Case (b) we can suppose, without loss of generality, that g is finite throughout e . Let us introduce the sets

$$e_n = e [g(x) \leq n] \quad (n = 1, 2, \dots)$$

Then $e = \bigcup_{n=1}^{\infty} e_n$ and the sets e_n form an increasing sequence, and we have

$$\int_{e_n} f d\mu \leq \int_{e_n} g d\mu \leq \int_E g d\mu$$

* As in analogous cases above, we suppose that the variation of the values of the function does not violate the property of measurability.

Corollary 2 of Theorem VII.2.2 implies

$$\int_{e_n} f d\mu \rightarrow \int_e f d\mu$$

and therefore

$$\int_e f d\mu \leq \int_E g d\mu \quad (7)$$

Thus, in both cases we obtain inequality (7). Now, passing to the supremum (over the sets e) in its left-hand side we arrive at formula (6).

Corollary. *If f is measurable and nonnegative on E and $f(x) \geq A > 0$ almost everywhere on E then*

$$\int_E f d\mu \geq A\mu E$$

Lemma VIII.2.5. *For any two nonnegative functions f and g measurable on E and any positive constant c we have*

$$\int_E (f + g) d\mu = \int_E f d\mu + \int_E g d\mu \quad (8)$$

and

$$\int_E cf d\mu = c \int_E f d\mu \quad (9)$$

In particular, if f and g are summable their sum and also their linear combination $\alpha f + \beta g$ with nonnegative coefficients α and β are also summable.

Proof. We can assume, without losing generality, that the functions f and g take finite values for all $x \in E$. Let us put $h = \alpha f + \beta g$ ($\alpha, \beta \geq 0$) and consider the sets

$$E_n = E [h(x) \leq n] \quad (n = 1, 2, \dots)$$

It is clear that $E = \bigcup_{n=1}^{\infty} E_n$, the sets E_n forming an increasing sequence.

Let us first suppose that $\mu E < +\infty$. Since the "linearity property" of the integral of a bounded measurable function over a set with finite measure was already proved (Theo-

rem VII.2.3) we have

$$\int_{E_n} h \, d\mu = \alpha \int_{E_n} f \, d\mu + \beta \int_{E_n} g \, d\mu \quad (10)$$

The finite additivity of the integral implies

$$\int_{E_n} h \, d\mu \xrightarrow{n \rightarrow \infty} \int_E h \, d\mu$$

(see the remark made after Lemma VIII.2.1). Analogous relations also hold for f and g . Therefore the passage to the limit in equality (10) results in

$$\int_E h \, d\mu = \alpha \int_E f \, d\mu + \beta \int_E g \, d\mu \quad (11)$$

In the general case when μE can be infinite ($\mu E = +\infty$) we represent E as $E = \bigcup_{n=1}^{\infty} E_n$ where the sets E_n form an increasing sequence and $\mu E_n < +\infty$. Using the linearity property of the integral over E_n which has already been proved and passing to the limit for $n \rightarrow \infty$ we again arrive at formula (11). Thus, both formulas (8) and (9) have been proved.

The Summability of the Powers of the Norm in R_n . Let us consider an important example. Let $\Delta = [-a, a]$ be a cube in R_n .^{*} For simplicity, we shall take $a = 1$; the general case can be treated analogously. Let us put

$$f(x) = \frac{1}{\|x\|^\alpha} \quad \text{where } \alpha > 0$$

for any $x \neq \theta$ belonging to R_n (here $\|\cdot\|$ is the Euclidean norm in R_n , that is $\|x\| = \sqrt{\sum_{i=1}^n \xi_i^2}$). We shall show that the function f is summable on Δ (with respect to the Lebesgue measure) if $\alpha < n$ and nonsummable if $\alpha \geq n$.^{**}

^{*} That is the parallelepiped specified by the inequalities $-a \leq \xi_i \leq a$ ($i = 1, 2, \dots, n$).

^{**} f is continuous on $R_n \setminus \{\theta\}$ and therefore measurable.

Taking the cube $\Delta_p = \left[-\frac{1}{p}, \frac{1}{p}\right]$ for any natural number p we can write

$$\Delta = \left\{ \bigcup_{p=1}^{\infty} (\Delta_p \setminus \Delta_{p+1}) \right\} \cup \{0\}$$

If $x \in \Delta_p \setminus \Delta_{p+1}$ then

$$\frac{1}{p+1} \leq \|x\| \leq \frac{\sqrt{n}}{p}$$

whence it follows that

$$\frac{p^\alpha}{n^{\alpha/2}} (\mu\Delta_p - \mu\Delta_{p+1}) \leq \int_{\Delta_p \setminus \Delta_{p+1}} f d\mu \leq (p+1)^\alpha (\mu\Delta_p - \mu\Delta_{p+1})$$

where μ is the Lebesgue measure in $\mathbf{R}_n$. But we have $\mu\Delta_p = \left(\frac{2}{p}\right)^n$, and hence

$$\frac{p^\alpha}{n^{\alpha/2}} 2^n \frac{(p+1)^n - p^n}{p^n (p+1)^n} \leq \int_{\Delta_p \setminus \Delta_{p+1}} f d\mu \leq (p+1)^\alpha 2^n \frac{(p+1)^n - p^n}{p^n (p+1)^n}$$

Consider the number series formed of the leftmost and rightmost members of these inequalities with summation with respect to p from 1 to ∞ . Their general terms are of the order of $\frac{1}{p^{n-\alpha+1}}$.* As is well known from the theory of series,

the series $\sum_{p=1}^{\infty} p^{-(n-\alpha+1)}$ is convergent for $\alpha < n$ and divergent for $\alpha \geq n$. Therefore $\int_{\Delta} f d\mu < +\infty$ for $\alpha < n$ and

$$\int_{\Delta} f d\mu = +\infty \text{ for } \alpha \geq n.$$

Analogous computations show that the function f is summable on the set $\mathbf{R}_n \setminus \Delta$ for $\alpha > n$ and nonsummable for $\alpha \leq n$.

§ 3. Generalization of the Simplest Properties of Integrals

In this section we shall show that almost all results of Sec. VII.2 concerning the integral of a bounded function

* The difference $(p+1)^n - p^n$ is of the order of p^{n-1} .

over a set with a finite measure are extended to the general case* treated in this chapter. As in the preceding section, all the functions under consideration are supposed to be finite almost everywhere.

Theorem VIII.3.1. *For a function f measurable on a set $E \subset X$ to be summable on that set it is necessary and sufficient that the function $|f|$ be summable on E .*

Proof. If f is summable on E then, by the definition of summability, the function $|f|$ is summable on each of the sets

$$E_1 = E [f(x) \geq 0] \text{ and } E_2 = E [f(x) < 0]$$

Therefore, by Lemma VIII.2.1, $|f|$ is summable on E .

Conversely, if $|f|$ is summable on E it is summable on E_1 and E_2 , that is, by definition, f is summable on E .

This theorem means that every summable function is automatically "absolutely summable". Therefore in the case of dimension 1 (i. e. in $\mathbf{R}_1$) there is a significant difference between the Lebesgue integral of an unbounded function or the integral over an infinite interval and the classical improper integral. As is known, in the classical theory of improper integrals a function defined on an interval or throughout the number line can be integrable in the improper sense and at the same time nonintegrable absolutely.** It can readily be proved that the classical improper integral in $\mathbf{R}_1$ of an absolutely integrable function coincides with the integral with respect to the Lebesgue measure (to this end one can use the remark made after Lemma VIII.2.1).

Theorem VIII.3.2. *If the integral $\int_E f d\mu$ exists then*

$$\left| \int_E f d\mu \right| \leq \int_E |f| d\mu \quad *** \quad (12)$$

* That is to the integral of a function with values of arbitrary sign over a set with arbitrary measure.

** In the space $\mathbf{R}_n$ with $n \geq 2$ there is no such difference between the Lebesgue integral and the classical improper integral. E.g. see B. M. Budak, S. V. Fomin, *Multiple Integrals, Field Theory and Series*, Moscow, Mir Publishers, 1973 (English edition).

*** We remind the reader that the right-hand member of formula (12) always makes sense.

Proof. From formula (2) it follows immediately that

$$\left| \int_E f d\mu \right| \leq \int_{E_1} |f| d\mu + \int_{E_2} |f| d\mu = \int_E |f| d\mu$$

Theorem VIII.3.3. *Let f be summable on a set E . Then, given any $\varepsilon > 0$, there is $\delta > 0$ such that*

$$\left| \int_{E'} f d\mu \right| < \varepsilon \text{ if } E' \subset E \text{ and } \mu E' < \delta$$

Proof. This theorem follows from the analogous Lemma VIII.2.2 proved for a nonnegative summable function and from inequality (12) since the summability of f implies that of $|f|$.

The property of the integral of a summable function expressed by Theorem VIII.3.3 is referred to as its *absolute continuity*.

The property of δ -additivity of the integral continues to hold in the general case and can be expressed by the following theorem.

Theorem VIII.3.4. *Let a set E be representable as a finite or countable union of disjoint measurable sets E_j ($E = \bigcup_j E_j$).*

Then:

(a) *if the integral $\int_E f d\mu$ exists each of the integrals $\int_{E_j} f d\mu$ also exists and*

$$\int_E f d\mu = \sum_j \int_{E_j} f d\mu \quad (13)$$

(b) *if f is summable on each E_j then f is summable on E if and only if the condition*

$$\sum_j \int_{E_j} |f| d\mu < +\infty \quad (14)$$

is fulfilled.

In particular, condition (14) always holds if E is broken up into a *finite* number of measurable sets E_j and, consequently, in this case the summability of f on the whole E follows from its summability on each set E_j .

Proof. (a) Let us resort to formula (4). Since the integral $\int_E f d\mu$ exists, at least one of the integrals $\int_E f_+ d\mu$ and $\int_E f_- d\mu$ is finite. For definiteness, let the latter integral be finite, that is let f_- be summable on E . Then f_- is summable on each E_j and, consequently, all the integrals $\int_{E_j} f d\mu$ exist. Each of them can have a finite value or be equal to $+\infty$. Similarly, $\int_E f d\mu$ is either finite or equal to $+\infty$. Therefore, on the basis of Lemma VIII.2.1, we obtain the following chain of equalities

$$\begin{aligned} \int_E f d\mu &= \int_E f_+ d\mu - \int_E f_- d\mu = \\ &= \sum_j \int_{E_j} f_+ d\mu - \sum_j \int_{E_j} f_- d\mu = \sum_j \int_{E_j} f d\mu \end{aligned}$$

each of which makes sense, which leads to formula (13).

(b) By Lemma VIII.2.1, we have

$$\int_E |f| d\mu = \sum_j \int_{E_j} |f| d\mu$$

Hence, condition (14) means that $|f|$ is summable on E , which, in its turn, is equivalent to the summability of the function f itself on the set E .

Corollary. If $E = \bigcup_{j=1}^{\infty} E_j$ where the sets E_j are measurable and disjoint, $\mu E_j < +\infty$ for all j , $f(x) = c_j$ on E_j ($j = 1, 2, \dots$; c_j are constants) and $\sum_{j=1}^{\infty} |c_j| \mu E_j < +\infty$ then f is summable on E and

$$\int_E f d\mu = \sum_{j=1}^{\infty} c_j \mu E_j$$

(this is a generalization of Corollary 1 of Theorem VII.2.2).

The corollary follows from parts (a) and (b) of the above theorem.

It should be noted that in condition (14) it is illegitimate to replace the integrals of $|f|$ by the integrals of f and require that (in the case of an infinite number of summands) the series $\sum_j \int_{E_j} f d\mu$ should converge. This can be confirmed by the following simple example.

Consider the function f defined in the interval $(0, 1]$ by the relations

$$f(x) = (-1)^n n \text{ for } \frac{1}{n+1} < x \leq \frac{1}{n}; \quad n = 1, 2, \dots$$

Then

$$\int_{1/(n+1)}^{1/n} f d\mu = \frac{(-1)^n}{n+1}$$

where μ is the Lebesgue measure, and the series $\sum_{n=1}^{\infty} \frac{(-1)^n}{n+1}$ is convergent. At the same time we have

$$\int_0^1 |f| d\mu = \sum_{n=1}^{\infty} \frac{1}{n+1} = +\infty *$$

As in Sec. VII.2, the σ -additivity of the integral proved above implies that if measurable sets E_p ($p = 1, 2, \dots$) form an increasing sequence, $E = \bigcup_{p=1}^{\infty} E_p$ and $\int_E f d\mu$ exists then

$$\int_E f d\mu = \lim_{p \rightarrow \infty} \int_{E_p} f d\mu \quad (15)$$

* Changing slightly this example we can also show that the condition of absolute convergence of the series of the integrals $\sum_j \int_{E_j} f d\mu$ in relation (14) is not sufficient either.

Theorem VIII.3.5. *If f and g are measurable on a set E , $f \sim g$ and at least one of the integrals $\int_E f d\mu$ and $\int_E g d\mu$ exists the other integral also exists and*

$$\int_E f d\mu = \int_E g d\mu$$

In particular, if f is summable on E then g is also summable.

Proof. If $f \sim g$ then $f_+ \sim g_+$ and $f_- \sim g_-$, and Lemma VIII.2.3 implies

$$\int_E f_+ d\mu = \int_E g_+ d\mu \text{ and } \int_E f_- d\mu = \int_E g_- d\mu$$

It follows immediately that the integrals $\int_E f d\mu$ and $\int_E g d\mu$ can only exist simultaneously and that they are equal.

Now the remark in the foregoing chapter made after the proof of Theorem VII.2.5 can be applied to the integrals of unbounded functions as well. In particular, it is possible to speak of the integral $\int_E f d\mu$ of a function defined almost everywhere on E without extending f to the remaining part of E .

One of the consequences of Theorem VIII.3.5 is that if $f(x) \geq 0$ almost everywhere on E and is measurable the integral of the function f makes sense and $\int_E f d\mu \geq 0$. If

$f \sim 0$ on E (f and E are supposed to be measurable) then $\int_E f d\mu = 0$. A converse of this assertion expressed for bounded functions by Theorem VII.2.6 continues to hold in the general case:

Theorem VIII.3.6. *If $f(x) \geq 0$ almost everywhere on E and $\int_E f d\mu = 0$ then $f \sim 0$.*

The proof of this theorem does not differ from that of the analogous Theorem VII.2.6 for bounded functions.

Theorem VIII.3.7. *If f and g are measurable on a set E , g is summable on E and $|f(x)| \leq |g(x)|$ almost everywhere on E the function f is also summable and*

$$\left| \int_E f d\mu \right| \leq \int_E g d\mu \quad (16)$$

Proof. By virtue of Theorem VIII.3.5, we can suppose without loss of generality that $g(x) \geq 0$ for all $x \in E$. Then Lemma VIII.2.4 indicates that the function $|f|$ is summable on E and, together with it, the function f is summable. Inequality (16) follows from the same lemma and Theorem VIII.3.2.

Theorem VIII.3.8. *For any two functions f and g summable on a set E and for any constant c , the functions $f \pm g$ and cf are also summable and*

$$\int_E (f \pm g) d\mu = \int_E f d\mu \pm \int_E g d\mu \quad (17)$$

$$\int_E cf d\mu = c \int_E f d\mu \quad (18)$$

Although the function $f \pm g$ is defined only almost everywhere and may not be defined on a subset of measure zero we can consider the integral of $f \pm g$ over the whole set E . The same remark applies to the product cf .*

Proof. Without loss of generality we can prove the theorem under the assumption that f and g are finite everywhere on E .

First of all we note that the inequality

$$|f(x) \pm g(x)| \leq |f(x)| + |g(x)|$$

of Lemma VIII.2.5 and Theorem VIII.3.7 imply the summability of the function $f \pm g$.

To begin with, let us derive formula (17) for the difference of the two nonnegative functions f and g . Putting $h(x) = f(x) - g(x)$ and taking into account that

$$h = h_+ - h_-, \quad f(x) \geq h_+(x) \quad \text{and} \quad g(x) \geq h_-(x)$$

(see Fig. 18 in Sec. VIII.2) we conclude that

$$f(x) - h_+(x) = g(x) - h_-(x) = k(x) \geq 0$$

* If $c = 0$ while $f(x) = \pm\infty$ the product cf makes no sense.

Consequently,

$$f(x) = h_+(x) + k(x) \quad \text{and} \quad g(x) = h_-(x) + k(x)$$

Whence, by Lemma VIII.2.4, the function $k(x)$ is summable.

According to Lemma VIII.2.5, we have

$$\int_E f d\mu = \int_E h_+ d\mu + \int_E k d\mu$$

and

$$\int_E g d\mu = \int_E h_- d\mu + \int_E k d\mu$$

Performing the term-by-term subtraction of these equalities (all the integrals involved have finite values) we obtain

$$\int_E f d\mu - \int_E g d\mu = \int_E h_+ d\mu - \int_E h_- d\mu = \int_E h d\mu = \int_E (f - g) d\mu$$

Now we shall prove formula (17) for the sum of the two functions f and g . By virtue of Lemma VIII.2.5, we have

$$\begin{aligned} \int_E f d\mu + \int_E g d\mu &= \int_E f_+ d\mu - \int_E f_- d\mu + \int_E g_+ d\mu - \int_E g_- d\mu = \\ &= \int_E (f_+ + g_+) d\mu - \int_E (f_- + g_-) d\mu \end{aligned} \quad (19)$$

Both sums in the parentheses are summable functions; their difference is equal to $f + g$ and, according to the above, the right-hand side of (19) is equal to $\int_E (f + g) d\mu$.

In the case $c = 0$ formula (18) is obvious. For $c \neq 0$ we put $l(x) = cf(x)$. If $c > 0$ then $l_+ = cf_+$, $l_- = cf_-$ and Lemma VIII.2.5 shows that

$$\int_E l_+ d\mu = c \int_E f_+ d\mu \quad \text{and} \quad \int_E l_- d\mu = c \int_E f_- d\mu$$

whence

$$\int_E l d\mu = c \int_E f d\mu \quad (20)$$

If $c < 0$ then $l_+ = |c| f_-$, $l_- = |c| f_+$, and equality (20) is obtained by means of analogous calculations.

Finally, formula (17) for the difference of two functions is obtained, without the requirement that the functions should be nonnegative, in the following way:

$$\begin{aligned} \int_E (f - g) d\mu &= \int_E [f + (-1)g] d\mu = \\ &= \int_E f d\mu + (-1) \int_E g d\mu = \int_E f d\mu - \int_E g d\mu \end{aligned}$$

Theorem VIII.3.9. *If f and g are summable on a set E and $f(x) \leq g(x)$ almost everywhere on E then*

$$\int_E f d\mu \leq \int_E g d\mu \quad (21)$$

Proof. We have $g(x) - f(x) \geq 0$ almost everywhere on E , and therefore

$$\int_E g d\mu - \int_E f d\mu = \int_E (g - f) d\mu \geq 0$$

which is equivalent to (21).*

We conclude this section with a brief discussion of the notion of an improper integral. This notion is not "absolute" and depends on what type of the integral is taken as the basic one, i.e. on what is meant by a "proper integral". In the classical mathematical analysis the basic notion of the integral is that of the Riemann integral of a bounded function over a finite interval, the integral of an unbounded function or the integral over an infinite interval being in this case considered improper. Here we shall take as a basic notion that of the integral of a summable function taken with respect to a measure and discuss a general approach to the notion of an improper integral.

Let $E \subset X$ be a measurable set represented in an arbitrary but fixed way as a decomposition into a system of measurable

* The last two theorems can also be proved if it is only required that the integrals $\int_E f d\mu$ and $\int_E g d\mu$ should make sense; but in this case we must introduce in Theorem VIII.3.8 the additional condition that the right-hand sides of formulas (17) and (18) also make sense.

sets E_n forming an increasing sequence: $E = \bigcup_{n=1}^{\infty} E_n$.

This representation of E as the union $\bigcup_{n=1}^{\infty} E_n$ will be denoted by R .

Definition. Let f be a measurable function defined on the set E . If f is summable on each of the sets E_n and if there exists a finite limit

$$\mathcal{J} = \lim_{n \rightarrow \infty} \int_{E_n} f d\mu$$

the number $\mathcal{J}$ is called the *improper* (or, more precisely, the *R-improper*) *integral* of the function f over the set E . As usual, in this case we say that the improper integral is *convergent* and denote it as

$$\int_E f d\mu \quad (\text{or, more precisely, } R\text{-}\int_E f d\mu)$$

The chosen representation R of the set E plays an essential role in this definition. Indeed, if R_1 and R_2 are two different representations of the set E as unions of increasing sequences of measurable sets it may happen that one of the integrals $R_1\text{-}\int_E f d\mu$ and $R_2\text{-}\int_E f d\mu$ is convergent while the other is divergent; it may also happen that both integrals are convergent but have different values.

If the function f is summable on the set E the σ -additivity of the integral with respect to measure implies the equality

$$\int_E f d\mu = \lim_{n \rightarrow \infty} \int_{E_n} f d\mu$$

Thus, in this case the *R-improper* integral is convergent for any representation R of the set E as a union of an increasing sequence of measurable sets and its value coincides with that of the integral with respect to the measure. If the function f is measurable and nonnegative the above equality holds even without the requirement that f should be summable. Therefore, in the latter case the convergence of

the R -improper integral of f over the set E for a concrete representation R implies the summability of f on E .

Thus, an improper integral of a nonnegative measurable function is convergent if and only if the function is summable. At the same time, in the classical theory of improper integrals an important role is played by the notion of an absolutely convergent improper integral. Following the definitions of that classical theory we should call the integral

$R\text{-}\int_E f d\mu$ absolutely convergent' if $R\text{-}\int_E |f| d\mu$ is convergent. However, as it has been shown, if the integral $R\text{-}\int_E |f| d\mu$ is convergent the function $|f|$ is summable

and then $|f|$ is also summable. Thus, we see that, under conditions assumed here, the notion of an absolutely convergent improper integral is of no interest; a wider class of functions for which the integral $\int_E f d\mu$ exists can only be

obtained if the *conditionally* (i.e. *not absolutely*) convergent integrals are included. As is well known, such integrals exist. For instance, if the integral over the interval $[0, +\infty)$ is understood as the limit of the sequence of integrals taken over an increasing sequence of intervals $[0, b_n]$ for $b_n \rightarrow +\infty$

the improper integral $\int_0^{+\infty} \frac{\sin x}{x} dx$ is convergent, but not

absolutely, i.e. conditionally, since the function $\left| \frac{\sin x}{x} \right|$ is nonsummable.

§ 4. Passage to the Limit under the Integral Sign

First of all we shall show that Lebesgue's Theorem VII.3.1 on the passage to the limit in the integral continues to hold under considerably wider assumptions. By the way, it should be noted that, while the "classical" condition of uniform convergence (which is rather restricting) is sufficient for the possibility of passing to the limit under the sign of the integral over a set of a finite measure, the situation is different for a set of infinite measure. For instance, let us

take the interval $(0, +\infty)$ and consider the functions

$$f_n(x) = \frac{n}{x^3} e^{-\frac{n}{2x^2}}, \quad n = 1, 2, \dots$$

All the functions are continuous and we readily find their greatest values $\frac{3\sqrt{3}}{\sqrt{n}} e^{-3/2}$ whence it follows that the sequence $\{f_n(x)\}$ uniformly converges to zero as $n \rightarrow \infty$. At the same time we have $\int_0^{+\infty} f_n dx = 1$ for all n , and consequently

the passage to the limit is illegitimate in this example.

Theorem VIII.4.1 (H. Lebesgue). *Let $\{f_k\}$ ($k = 1, 2, \dots$) be a sequence of functions summable on a set $E \subset X$ which converges in measure to a function f ($f_k \Rightarrow f$). If there exists a nonnegative summable function φ such that $|f_k(x)| \leq \varphi(x)$ for each k almost everywhere on E , the function f is also summable and*

$$\lim \int_E f_k d\mu = \int_E f d\mu^* \quad (22)$$

Proof. Riesz' theorem VI.5.3 indicates that there exists a subsequence $\{f_{k_i}\}$ convergent to f almost everywhere on E . Therefore it becomes clear that $|f(x)| \leq \varphi(x)$ almost everywhere and hence, by Theorem VIII.3.7, the function f is summable.

Let us take an arbitrary $\varepsilon > 0$. From the definition of the integral it follows that there exists a set $E' \subset E$ with $\mu E' < +\infty$ such that†

$$\int_{E \setminus E'} \varphi d\mu < \frac{\varepsilon}{6} \quad (23)$$

(if $\mu E < +\infty$ we can take $E' = E$). Furthermore, by Theorem VIII.3.3, there exists $\delta > 0$ such that $\int_{E''} \varphi d\mu < \frac{\varepsilon}{6}$

* Since the integrand can be replaced by any equivalent function without changing the value of the integral we can suppose that the functions f_k and f assume finite values at all points. By the way, this remark is not very important since the convergence in measure can also be considered without this restriction (cf. Sec. VI.5).

if $E'' \subset E$ and $\mu E'' < \delta$. Let us choose $\eta > 0$ such that $\eta \mu E' \leq \frac{\varepsilon}{3}$. The definition of convergence in measure implies that there is K such that for $k \geq K$ we have

$$\mu E' [|f_k(x) - f(x)| \geq \eta] < \delta$$

Now let us break up the set E , for each k , into the three subsets

$$E_1 = E \setminus E', \quad E_2 = E' [|f_k(x) - f(x)| \geq \eta]$$

$$\text{and } E_3 = E' [|f_k(x) - f(x)| < \eta]$$

Since $|f_k(x) - f(x)| \leq 2\varphi(x)$ almost everywhere on E , inequality (23) and Theorem VIII.3.7 show that we have

$$\left| \int_{E_1} (f_k - f) d\mu \right| < \frac{\varepsilon}{3}$$

for all $k = 1, 2, \dots$. If $k \geq K$ then $\mu E_2 < \delta$ and, according to the choice of δ ,

$$\left| \int_{E_2} (f_k - f) d\mu \right| < \frac{\varepsilon}{3}$$

Finally, for all k we have

$$\left| \int_{E_3} (f_k - f) d\mu \right| \leq \eta \mu E_3 \leq \eta \mu E' \leq \frac{\varepsilon}{3}$$

Thus,

$$\left| \int_E f_k d\mu - \int_E f d\mu \right| < \varepsilon$$

for $k \geq K$, which proves equality (22).

Theorem VII.3.1 is a special case of Theorem VIII.4.1 since in the former $\mu E < +\infty$ and consequently the constant M bounding the functions f_k in this theorem can be regarded as a summable function on E and thus can play the role of the function φ .

Remark. The proof of Lebesgue's theorem indicates that the theorem continues to hold when in its statement the convergence in measure is replaced by the convergence almost everywhere (to a measurable function). Indeed, in the

proof of Lebesgue's theorem it is sufficient to use the convergence in measure on a subset E' with $\mu E' < +\infty$, and under the last condition the convergence almost everywhere implies the convergence in measure (see Theorem VI.5.2).

There is a stronger theorem for monotone functional sequences proved by the Italian mathematician B. Levi (1875-1961). We shall state it here for a monotone sequence of nonnegative functions.

Theorem VIII.4.2. *If there is a nondecreasing sequence of functions $f_k(x) \geq 0$ ($k = 1, 2, \dots$; $f_k(x) \leq f_{k+1}(x)$ for all k) measurable on a set E and $f_k(x) \rightarrow f(x)$ for all $x \in E$ where f is finite almost everywhere on E there holds equality (22):*

$$\lim \int_E f_k d\mu = \int_E f d\mu$$

Proof. It follows from Lemma VIII.2.4 that the integrals $\int_E f_k d\mu$ form a nondecreasing number sequence and $\int_E f_k d\mu \leq \int_E f d\mu$ for all k , whence

$$\lim \int_E f_k d\mu \leq \int_E f d\mu$$

To obtain the opposite inequality* let us take an arbitrary measurable set $e \subset E$ with $\mu e < +\infty$ on which f is bounded. By Lebesgue's theorem,

$$\int_e f d\mu = \lim \int_e f_k d\mu \leq \lim \int_E f_k d\mu$$

Passing to the supremum in the left-hand side we see that

$$\int_E f d\mu \leq \lim \int_E f_k d\mu$$

which completes the proof of equality (22).

Remark. Let us show that in Levi's theorem it is possible to discard the requirement that all the functions should be finite almost everywhere.

* The opposite inequality also follows from Theorem VIII.4.3 given below.

Suppose that all the other conditions of Levi's theorem hold and $\mu E [f(x) = +\infty] > 0$. Then, by definition, we have $\int_E f d\mu = +\infty$, and it remains to prove that $\lim \int_E f_k d\mu = +\infty$, that is to show that the integrals $\int_E f_k d\mu$ are not uniformly bounded. Assume the contrary, i.e. let $\int_E f_k d\mu \leq M$ for all k . If C is arbitrary positive constant we can write, by the foregoing inequality,

$$\mu E [f_k(x) > C] \leq \frac{M}{C}$$

$$(k = 1, 2, \dots)$$

But

$$E [f(x) > C] = \bigcup_{k=1}^{\infty} E [f_k(x) > C]$$

and the sets on the right-hand side form an increasing sequence whence

$$\mu E [f(x) > C] \leq \frac{M}{C}$$

This shows that

$$\mu E [f(x) = +\infty] \leq \frac{M}{C}$$

for any C , that is $\mu E [f(x) = +\infty] = 0$, and we thus arrive at a contradiction.

Let us introduce another useful notion. If f is a nonnegative function measurable on a set E we define, for every natural number m and every $x \in E$, the *truncated function*

$$f_m(x) = \begin{cases} f(x) & \text{if } f(x) \leq m \\ m & \text{if } f(x) > m \end{cases}$$

(see Fig. 19). In other words, all the values of $f(x)$ exceeding m are replaced by the number m . From Lemma VI.3.1 and the obvious equality

$$f_m(x) = \min [f(x), m]$$

it follows that all f_m ($m = 1, 2, \dots$) are bounded and measurable on E .

Let us indicate some consequences of Levi's theorem and of the above remark.

(a) For any nonnegative measurable function $f(x)$ and the corresponding sequence of the truncated functions $f_m(x)$ ($m = 1, 2, \dots$) it is allowable to pass to the limit under the sign of integral.

(b) If, in addition to the conditions of Levi's theorem, we suppose that the integrals $\int_E f_k d\mu$, $k = 1, 2, \dots$ are uni-

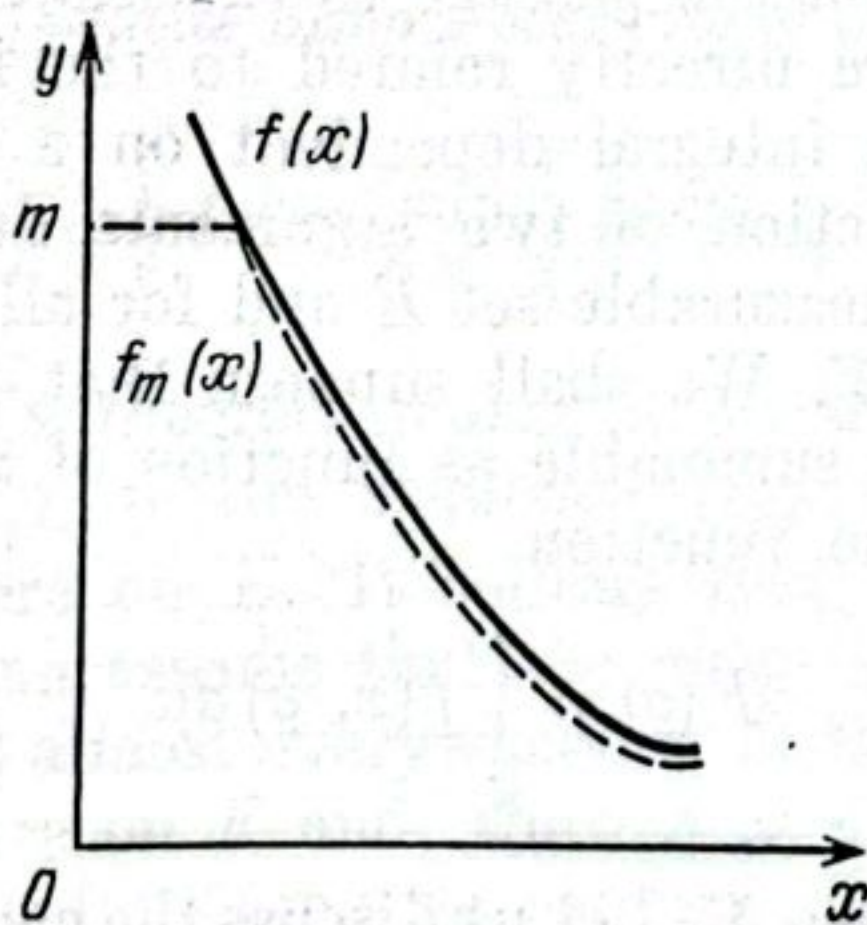


Fig. 19

formly bounded this will imply that the limit function f is finite almost everywhere.

(c) If $\{u_k(x)\}$ is a sequence of nonnegative functions ($u_k(x) \geq 0$; $k = 1, 2, \dots$) summable on a set E and $\sum_{k=1}^{\infty} \int_E u_k d\mu < +\infty$, the function $s(x) = \sum_{k=1}^{\infty} u_k(x)$ is finite almost everywhere on E and

$$\int_E s d\mu = \sum_{k=1}^{\infty} \int_E u_k d\mu$$

Proof. Let us put $s_m(x) = \sum_{k=1}^m u_k(x)$. Then the sequence $s_m(x)$ is nondecreasing and $s_m(x) \rightarrow s(x)$, the integrals

$\int_E s_m d\mu$ ($m = 1, 2, \dots$) being uniformly bounded. By the foregoing proposition, this implies that s is finite almost everywhere on E , and from Theorem VIII.4.2 it follows that

$$\int_E s d\mu = \lim \int_E s_m d\mu = \lim \sum_{k=1}^m \int_E u_k d\mu = \sum_{k=1}^{\infty} \int_E u_k d\mu$$

The theorems on the passage to the limit under the sign of integration are directly related to the investigation of continuity of an integral dependent on a parameter. Let $f(x, y)$ be a function of two arguments defined for all x belonging to a measurable set E and for all y belonging to a metric space Y . We shall suppose that for every $y \in Y$ the function f is summable as function of x on the set E . This specifies the function

$$F(y) = \int_E f(x, y) d\mu$$

defined on the space Y . Let us discuss the conditions guaranteeing the continuity of the function F at a given point $y_0 \in Y$, i.e. the fulfilment of the relation

$$\int_E f(x, y) d\mu \xrightarrow{y \rightarrow y_0} \int_E f(x, y_0) d\mu$$

This relation means that, for any sequence $\{y_n\}$ convergent to y_0 ($y_n \rightarrow y_0$), the relation

$$\int_E f(x, y_n) d\mu \rightarrow \int_E f(x, y_0) d\mu$$

should hold. Now we see that the theorems on the passage to the limit under the integral sign provide sufficient conditions for the latter relation to be fulfilled. For instance, if the function f is continuous with respect to y at the point y_0 for every $x \in E$ (or for almost all $x \in E$) and possesses a summable majorant, that is

$$|f(x, y)| \leq \varphi(x)$$

for all $x \in E$, $y \in E$ (or at least for all y belonging to a neighbourhood of y_0) where φ is summable on E , the Lebesgue theorem on the passage to the limit in the integral is applicable and therefore the function F is continuous at the point y_0 .

We shall also present a theorem proved by P. Fatou (1878-1929), a French mathematician. This theorem establishes a relation slightly weaker than equality (22).

Theorem VIII.4.3. *Let a sequence of functions $f_k(x) \geq 0$, $k = 1, 2, \dots$, measurable and finite almost everywhere, be defined on a set E . If this sequence converges in measure to a function f ($f_k \Rightarrow f$) finite almost everywhere on E then*

$$\int_E f d\mu \leq \sup_k \int_E f_k d\mu \quad (24)$$

Proof.* Riesz's theorem makes it possible to choose a subsequence $\{f_{k_i}\}$ of the sequence $\{f_k\}$ convergent to f almost everywhere on E . It follows that, without loss of generality, we can assume that the given sequence $\{f_k(x)\}$ converges to $f(x)$ almost everywhere on E . Besides, $f(x) \geq 0$ almost everywhere on E and, since it is allowable to pass to an equivalent function, we can assume that $f(x) \geq 0$ for all $x \in E$.

Let us introduce the functions

$$g_k(x) = \min [f_k(x), f(x)] \quad (k = 1, 2, \dots)$$

By Lemma VI.3.1 they are measurable, $g_k(x) \leq f(x)$ for $k = 1, 2, \dots$ and, obviously,

$$g_k(x) \rightarrow f(x) \text{ for almost all } x \in E$$

Let us consider the following two possible cases. Suppose first that f is summable on E . Then Theorem VIII.4.1 applies (see the remark made after this theorem), and consequently

$$\int_E g_k d\mu \rightarrow \int_E f d\mu$$

* The proof given here was communicated to the author by G. I. Natanson.

At the same time we have $\int_E f_k d\mu \geq \int_E g_k d\mu$ for all k and therefore

$$\sup_k \int_E f_k d\mu \geq \int_E g_k d\mu$$

Now passing to the limit we directly arrive at (24).

Now let us suppose that f is nonsummable on E . According to what has already been proved, for any arbitrary set $e \subset E$ with a finite measure on which f is bounded we have

$$\sup_k \int_E f_k d\mu \geq \sup_k \int_e f_k d\mu \geq \int_e f d\mu$$

Passing to the supremum in the right-hand side we again obtain (24).

Remark. The proof of Fatou's theorem shows directly that it also applies to the case when $f_k(x) \rightarrow f(x)$ almost everywhere on E and f is measurable. The proof was in fact carried out for this particular case. In addition, let us show that in the case of the convergence almost everywhere the requirement that the functions f_k and f should be finite almost everywhere can be discarded in the formulation of Fatou's theorem.

To this end we take the truncated functions $f_k^{(p)}$ and $f^{(p)}$ corresponding, respectively, to the functions f_k and f .^{*} It is evident that $f_k^{(p)}(x) \xrightarrow[k \rightarrow \infty]{} f^{(p)}(x)$ almost everywhere on E . By Fatou's theorem (which has already been proved for finite functions), we have

$$\int_E f^{(p)} d\mu \leq \sup_k \int_E f_k^{(p)} d\mu$$

and therefore

$$\int_E f^{(p)} d\mu \leq \sup_k \int_E f_k d\mu$$

^{*} Here p is an integer and $f_k^{(p)}(x) = \min[f_k(x), p]$, $f^{(p)}(x) = \min[f(x), p]$.

On the other hand, according to Corollary (a) of Levi's theorem,

$$\int_E f^{(p)} d\mu \rightarrow \int_E f d\mu$$

and we thus obtain inequality (24).

Levi's theorem and the theorem on the σ -additivity of the integral indicate another way of passing from the integral of a bounded function over a set with finite measure to the integral of an arbitrary nonnegative measurable function f . First of all, if the function f is defined on a set E with $\mu E < +\infty$ the integral $\int_E f d\mu$ can be defined as the limit of the sequence of the integrals of the corresponding truncated functions. Next, if $\mu E = +\infty$ we can choose an increasing sequence of sets E_p with finite measure such that $E = \bigcup_{p=1}^{\infty} E_p$ and define the integral of f over E by means of formula (15):

$$\int_E f d\mu = \lim \int_{E_p} f d\mu$$

It follows from Theorems VIII.4.2 and VIII.3.4 that the definition thus stated is equivalent to the one given in Sec. VIII.1.

The passage to the limit in the integral is closely related to the notion of an equicontinuous family of integrals.

Definition. Let there be functions f_k ($k = 1, 2, \dots$) summable on a set E . The family of the integrals of these functions will be called *absolutely equicontinuous* if, given an arbitrary $\varepsilon > 0$, there is $\delta > 0$ such that

$$\int_e |f_k| d\mu < \varepsilon \quad \text{for all } k$$

for any measurable set $e \subset E$ with $\mu e < \delta$.

This definition can of course be stated not only for functional sequence but also for an arbitrary collection of functions.

It is readily seen that if the functions f_k possess a summable majorant φ (this means that $|f_k(x)| \leq \varphi(x)$ al-

most everywhere on E for any k) the family of the integrals of f_k is absolutely equicontinuous.* It is the property that was in fact used in the proof of Theorem VIII.4.1. A more thorough analysis shows that Theorem VIII.4.1 can be (partially) strengthened:

Theorem VIII.4.4 (G. Vitali). *If f_k ($k = 1, 2, \dots$) are summable functions on a set E with $\mu E < +\infty$, f is a measurable function finite almost everywhere on E , $f_k \Rightarrow f$ and the family of the integrals of f_k is absolutely equicontinuous, the function f is also summable on E and equality (22) holds.*

It should be noted that if $\mu E = +\infty$ the absolute equicontinuity of the integrals is no longer sufficient for the possibility of passing to the limit under the sign of integration, and therefore Lebesgue's theorem (with all its assertions) is not a consequence of Theorem VIII.4.4.

Proof. Take an arbitrary $\varepsilon > 0$ and put $\eta = \frac{\varepsilon}{\mu E}$. Let $\delta > 0$ be the number, indicated in the definition of the absolute equicontinuity, corresponding to the chosen ε . From the inequality

$$||f_k(x)| - |f(x)|| \leq |f_k(x) - f(x)|$$

it follows that $|f_k| \Rightarrow |f|$ on E . Therefore, if $e \subset E$ and $\mu e < \delta$ Fatou's theorem implies

$$\int_e |f| d\mu \leq \sup_k \int_e |f_k| d\mu \leq \varepsilon$$

Now let us consider the sets

$$A_k = E[|f_k(x) - f(x)| < \eta], \quad B_k = E \setminus A_k$$

If k is sufficiently large we have $\mu B_k < \delta$ and therefore

$$\int_{B_k} |f_k - f| d\mu \leq \int_{B_k} |f_k| d\mu + \int_{B_k} |f| d\mu < 2\varepsilon$$

On the other hand,

$$\int_{A_k} |f_k - f| d\mu \leq \eta \cdot \mu A_k \leq \eta \cdot \mu E = \varepsilon$$

* The converse of this assertion is not true; this can be confirmed by some simple examples.

Consequently,

$$\int_E |f_k - f| d\mu < 3\varepsilon$$

In particular, it follows that the function $f_k - f$ is summable on E ; therefore the function $f = f_k - (f_k - f)$ is also summable and

$$\int_E f_k d\mu \rightarrow \int_E f d\mu$$

To demonstrate the application of the last theorem we shall prove the continuity of the *integral of the potential type*.

Let E be a bounded Lebesgue measurable set in the space R_n and $B(x, y)$ be a bounded continuous function defined for $x, y \in E$ and satisfying the condition $|B(x, y)| \leq M$. Consider the integral

$$I(y) = \int_E \frac{B(x, y)}{\|x - y\|^\alpha} d\mu_x \quad (25)$$

where $0 < \alpha < n$, $\mu = \mu_x$ is Lebesgue's measure and the subscript x indicates that the integration is performed "with respect to x ", y playing the role of a parameter. The norm $\|x - y\|$ is understood as an ordinary Euclidean norm in R_n .

Since E is bounded there exists a parallelepiped $\Delta \subset R_n$ such that $x - y \in \Delta$ for any $x, y \in E$. The function $1/\|u\|^\alpha$ is summable on Δ (see Sec. VIII.2) and therefore for any $\varepsilon > 0$ there is $\delta > 0$ such that

$$\int_e \frac{d\mu}{\|u\|^\alpha} < \varepsilon \quad \text{if } e \subset \Delta \quad \text{and } \mu e < \delta$$

Now let us show that the family of the integrals of the functions $B(x, y)/\|x - y\|^\alpha$ (taken "with respect to x ") is absolutely equicontinuous (with respect to y). Indeed, if $e \subset E$ is a measurable set then, for any $y \in E$, the set e' obtained as e is translated by the vector $-y$ (i.e. $e' = \{x - y \mid x \in e\}$) is contained in Δ and has the same

measure as e . Therefore, if $\mu e < \delta$ we have

$$\int_e \frac{|B(x, y)|}{\|x - y\|^\alpha} d\mu_x \leq M \int_e \frac{d\mu_x}{\|x - y\|^\alpha} = M \int_{e'} \frac{d\mu}{\|u\|^\alpha} < M\epsilon^*$$

which proves the absolute equicontinuity of the integrals of the given functions. The continuity of the function $B(x, y)$ implies that if $y_k \rightarrow y_0$ where $y_k, y_0 \in E$ then, for $x \neq y_0$, there must be

$$\frac{B(x, y_k)}{\|x - y_k\|^\alpha} \rightarrow \frac{B(x, y_0)}{\|x - y_0\|^\alpha}$$

The convergence for all $x \neq y_0$ ($x \in E$) implies the convergence in measure, and hence, by Theorem VIII.4.4,

$$I(y_k) \rightarrow I(y_0)$$

that is integral (25) is a continuous function of y .

The solution of the problem of differentiability of an integral with respect to a parameter is also connected with the theorems on passing to the limit under the sign of integration. Let $f(x, y)$ be summable as function of x on a set E for each y belonging to an interval $(c, d) \subset \mathbf{R}_1$. Putting

$$F(y) = \int_E f(x, y) d\mu$$

we can state the following theorem.

Theorem VIII.4.5. *Let*

(i) *for any $x \in E$ and $y \in (c, d)$ there exist the partial derivative f'_y ;*

(ii) *there exist a function φ summable on E such that*

$$|f'_y(x, y)| \leq \varphi(x)$$

for all $x \in E$ and $y \in (c, d)$.

Then the function F is differentiable at every point $y_0 \in (c, d)$ and

$$F'(y_0) = \int_E f'_y(x, y_0) d\mu$$

Moreover, if f'_y is continuous in y at the point y_0 for every $x \in E$ the derivative F' is also continuous at that point.

* The middle equality obtained by change of variable $x - y = u$ is quite obvious.

Proof. We have

$$f'_y(x, y) = \lim_{k \rightarrow 0} \frac{f(x, y+k) - f(x, y)}{k}$$

and consequently f'_y can be represented as the limit of a sequence of measurable functions for each y . Therefore the function f'_y is measurable as function of x for each y . Furthermore, we have

$$\frac{F(y_0+k) - F(y_0)}{k} = \frac{1}{k} \int_E [f(x, y_0+k) - f(x, y_0)] d\mu$$

Since

$$\frac{1}{k} [f(x, y_0+k) - f(x, y_0)] \xrightarrow{k \rightarrow 0} f'_y(x, y_0)$$

and

$$\left| \frac{1}{k} [f(x, y_0+k) - f(x, y_0)] \right| = |f'_y(x, y_0 + \theta k)| \leq \varphi(x)$$

the integral on the right-hand side of the desired equality satisfies the conditions of the Lebesgue theorem which implies

$$\frac{1}{k} \int_E [f(x, y_0+k) - f(x, y_0)] d\mu \xrightarrow{k \rightarrow 0} \int_E f'_y(x, y_0) d\mu$$

The theorem has thus been proved. The Lebesgue theorem also shows that the derivative F' is continuous provided that f'_y is continuous in y .

§ 5. The Space L of Summable Functions

In this section we shall consider the collection of all functions defined and summable on a fixed set $E \subset X$ and introduce a norm in it to obtain a normed space (see Sec. III.6). Let us denote this collection by L . To indicate, if necessary, the domain of definition of the functions we shall also write L_E . As in the definition of the space S (Sec. VII.4), we shall identify in L functions equivalent to each other; this will allow us to regard every element of the set L as represented by a function assuming finite values everywhere. This convention facilitates the arithmetical operations on functions belonging to L .

Theorem VIII.3.8 shows that $\mathcal{L}$ is a linear space. If we take the sum $f + g$ of two functions f and g belonging to $\mathcal{L}$ and replace each summand by an equivalent function contained in $\mathcal{L}$, that is consider the sum $f_1 + g_1$ where $f_1 \sim f$ and $g_1 \sim g$, the latter sum will be equivalent to the former. This means that these sums coincide as elements of the set $\mathcal{L}$. A similar remark applies to the product of a function by a number. The zero element θ of the set $\mathcal{L}$ is represented not only by the function identically equal to 0 but also by every function f equivalent to 0.

To introduce a norm in the set $\mathcal{L}$ we put

$$\|f\| = \int_E |f| d\mu$$

for any function $f \in \mathcal{L}$.

It is clear that $\|f\| \geq 0$ for any $f \in \mathcal{L}$, and from Theorem VIII.3.6 it follows that $\|f\| = 0$ if and only if $f \sim 0$, that is if $f = \theta$ as element of the set $\mathcal{L}$. Theorem VIII.3.8 indicates that $\|cf\| = |c| \|f\|$ where c is a constant. Finally, Theorems VIII.3.8 and VIII.3.9 imply the triangle inequality for the norm:

$$\|f + g\| = \int_E |f + g| d\mu \leq \int_E |f| d\mu + \int_E |g| d\mu = \|f\| + \|g\|$$

Hence, $\mathcal{L}$ is a *normed space*.

The relation $f_k \rightarrow f$ in $\mathcal{L}$ is understood as the convergence in norm and means that

$$\int_E |f_k - f| d\mu \rightarrow 0$$

Such a sequence f_k is said to be *convergent to f in the mean of order one*. In this section we shall often simply say "convergent in the mean".

Theorem VIII.5.1. *If $f_k \rightarrow f$ in norm in the space $\mathcal{L}$ then $f_k \Rightarrow f$ (that is, the convergence in the mean implies the convergence in measure to the same limit function).*

Proof. Suppose that the sequence $\{f_k\}$ is not convergent in measure to f . This means that for some $\varepsilon > 0$ there is $\delta > 0$ such that

$$\mu E [|f_k(x) - f(x)| \geq \varepsilon] \geq \delta \quad (26)$$

for an infinite number of values of the index k : $k = k_1, k_2, \dots, k_i, \dots$. Denoting by e_k the set on the left-hand side of (26) we can write

$$\int_E |f_{k_i} - f| d\mu \geq \int_{e_k} |f_{k_i} - f| d\mu \geq \delta \varepsilon$$

Thus, the subsequence $\{f_{k_i}\}$ cannot converge in the mean to f . The theorem has been proved.

The converse assertion is not true: a sequence of summable functions may be convergent in measure to a summable function and at the same time may not be convergent in the mean. This fact is confirmed by the example given at the beginning of Sec. VII.3. In this example the interval $(0, 1)$ plays the role of E , $f_k(x) \rightarrow 0$ throughout that interval and the functions f_k are summable. At the same time $f_k \not\rightarrow 0$ with respect to the norm of the space $L_{(0,1)}$ since $\|f_k\| = 1$ for all k .

Theorem VIII.5.2. L is a Banach space.

Proof. It is only necessary to verify the completeness of the space L . Let $\{f_k\}$ be a Cauchy sequence in the space L , that is $\|f_k - f_l\| \rightarrow 0$ for $k, l \rightarrow \infty$. By analogy with the proof of the foregoing theorem we readily show that, given any $\varepsilon > 0$, there must be

$$\mu E[|f_k(x) - f_l(x)| \geq \varepsilon] \xrightarrow[k, l \rightarrow \infty]{} 0$$

According to Lemma VI.5.1 it is possible to choose a subsequence $\{f_{k_i}\}$ of $\{f_k\}$ which converges almost everywhere on E to a limit function f with finite values.

Let us show that f is summable. Indeed, taking an arbitrary $\eta > 0$ we can write, for all sufficiently large k and l ($k, l \geq K$), the inequality

$$\|f_k - f_l\| < \eta$$

In particular, $\|f_{k_i} - f_{k_j}\| < \eta$, that is

$$\int_E |f_{k_i} - f_{k_j}| d\mu < \eta$$

for all sufficiently large i and j (such that $k_i, k_j \geq K$). For a fixed i we have

$$|f_{k_i}(x) - f_{k_j}(x)| \xrightarrow{j \rightarrow \infty} |f_{k_i}(x) - f(x)|$$

almost everywhere on E , and, by Fatou's Theorem VIII.4.3,

$$\int_E |f_{k_i} - f| d\mu \leq \sup_{k_j \geq K} \int_E |f_{k_i} - f_{k_j}| d\mu \leq \eta$$

Consequently, the function $f_{k_i} - f$ is summable on E , and then the function

$$f = (f - f_{k_i}) + f_{k_i}$$

is summable as well. Incidentally we have proved that $\|f_{k_i} - f\| \leq \eta$ for all sufficiently large i . Now, by the arbitrariness of η , we conclude that $f_{k_i} \rightarrow f$ with respect to the norm of the space $\mathbf{L}$. Finally, from Lemma III.4.1 it follows that $f_k \rightarrow f$, and thus the completeness of $\mathbf{L}$ has been proved.

Now let us dwell on the case when the space $\mathbf{L}$ consists of the summable functions defined on a set $E \subset \mathbf{R}_n$, μ being the Lebesgue measure. We shall prove that such a space $\mathbf{L}$ is separable. Let E_p denote, for any natural number p , the intersection of the set E with the closed parallelepiped

$$\Delta_p^* = [-p, p; -p, p; \dots; -p, p]$$

and let H_p be the set of all functions ψ defined on E such that $\psi(x)$ is equal, on E_p , to an algebraic polynomial (in n variables) with rational coefficients and $\psi(x) = 0$ on $E \setminus E_p$. Every such function is summable on E and, consequently, $H_p \subset \mathbf{L}$. All the sets H_p are countable and therefore the set $H = \bigcup_{p=1}^{\infty} H_p$ is also countable. Let us show that H is everywhere dense in $\mathbf{L}$.*

Let $f \in \mathbf{L}$. Choosing an arbitrary $\varepsilon > 0$ we can take p such that

$$\int_{E \setminus E_p} |f| d\mu = \int_E |f| d\mu - \int_{E_p} |f| d\mu < \frac{\varepsilon}{4} \quad (27)$$

* If the set E is bounded we can take as H the set of the polynomials (with rational coefficients) itself.

(see formula (15)). Next we can take a measurable set $e \subset E_p$ such that f is bounded on e [$|f(x)| < M$] and

$$\int_{E_p \setminus e} |f| d\mu = \int_{E_p} |f| d\mu - \int_e |f| d\mu < \frac{\varepsilon}{4}$$

Let us put

$$g(x) = \begin{cases} f(x) & \text{for } x \in e \\ 0 & \text{for } x \in E_p \setminus e \end{cases}$$

Then

$$\int_{E_p} |f - g| d\mu = \int_{E_p \setminus e} |f| d\mu < \frac{\varepsilon}{4} \quad (28)$$

By Luzin's theorem, there is a function φ continuous throughout R_n satisfying the condition $|\varphi(x)| \leq M$ (see the remark made after Theorem VI.6.4) such that

$$\mu E_p [g(x) \neq \varphi(x)] < \frac{\varepsilon}{8M}$$

Since $|g(x) - \varphi(x)| < 2M$ throughout E_p we have

$$\int_{E_p} |g - \varphi| d\mu < 2M \frac{\varepsilon}{8M} = \frac{\varepsilon}{4} \quad (29)$$

Finally, let us choose an algebraic polynomial P (in the n variables $\xi_1, \xi_2, \dots, \xi_n$) with rational coefficients such that

$$|\varphi(x) - P(x)| < \frac{\varepsilon}{4\mu\Delta_p^*} \quad \text{for all } x \in \Delta_p^*$$

and put

$$\psi(x) = \begin{cases} P(x) & \text{for } x \in E_p \\ 0 & \text{for } x \in E \setminus E_p \end{cases}$$

Then

$$\int_{E_p} |\varphi - \psi| d\mu \leq \frac{\varepsilon}{4\mu\Delta_p^*} \cdot \mu E_p \leq \frac{\varepsilon}{4} \quad (30)$$

From the definition of ψ it follows that $\psi \in H_p \subset H$ and that

$$\|f - \psi\| = \int_{E_p} |f - \psi| d\mu + \int_{E \setminus E_p} |f| d\mu$$

Therefore (27), (28), (29) and (30) imply $\|f - \psi\| < \varepsilon$. We have thus proved that the set H is everywhere dense in L .

As was mentioned in Sec. III.4, the space C_L of all continuous functions defined on an interval $[a, b]$ is not complete.* Here we shall prove this fact. It follows from the above proof that the set of all continuous functions defined in the interval $[a, b]$ is everywhere dense in $L_{[a, b]}$. Let us choose a function belonging to $L_{[a, b]}$ which is not equivalent to any continuous function. For instance, such is the function

$$f(x) = \begin{cases} 1 & \text{for } a \leq x \leq c \\ 0 & \text{for } c < x \leq b \end{cases}$$

where $a < c < b$. There exists a sequence of continuous functions f_k converging in norm to f in the space $L_{[a, b]}$ **. Then $\{f_k\}$ is a Cauchy sequence in the sense of the convergence in the mean which is exactly the convergence in the sense of the metric in C_L ***. On the other hand, the uniqueness of the limit in L implies that the sequence $\{f_k\}$ cannot be convergent in the mean to any continuous function, that is it cannot have a limit in C_L .

Let us come back to the general case when X can be an arbitrary measure space and discuss the properties of linear functionals defined in the space L_E .

* Similarly, the space of all continuous functions of several arguments (defined, for instance, in a closed parallelepiped with finite edges) metrized in a similar way is not complete either.

** For instance, we can put

$$f_k(x) = \begin{cases} 1 & \text{for } a \leq x \leq c - \frac{c-a}{k} \\ 0 & \text{for } c \leq x \leq b \end{cases}$$

and define $f_k(x)$ by means of linear interpolation over the interval $(c - \frac{c-a}{k}, c)$.

*** If we equip C_L with norm using the same formula as in the case of the space L then C_L becomes a normed space in which the distance is defined as before:

$$\rho(f, g) = \|f - g\| = \int_a^b |f - g| dx$$

We shall say that a function φ defined on a set E is *bounded almost everywhere* on this set if there is a constant C such that $|\varphi(x)| \leq C$ almost everywhere on E . Among the constants C satisfying this condition there exists the least one. Indeed, let C_0 be equal to the greatest lower bound of such constants. Assuming, for simplicity, that the function φ is measurable and putting

$$e_m = E \left[|\varphi(x)| > C_0 + \frac{1}{m} \right]$$

for every natural number m we have $\mu e_m = 0$. If $e = \bigcup_{m=1}^{\infty} e_m$ then $\mu e = 0$ as well. It is clear that $|\varphi(x)| \leq C_0$ for all $x \in E \setminus e$ and consequently $|\varphi(x)| \leq C_0$ almost everywhere on E , and from the definition of C_0 it follows that it is the least of the constants C . ■

The least constant C_0 is called the *essential supremum* of the function $|\varphi|$ and is denoted as

$$\text{vrai sup}_{x \in E} |\varphi(x)|^*$$

Now let φ be an arbitrary function bounded almost everywhere and measurable on a set E and let $C_0 = \text{vrai sup}_{x \in E} |\varphi(x)|$. Then the product $f\varphi$ is measurable for any $f \in L$ and $|f(x)\varphi(x)| \leq C_0 |f(x)|$ almost everywhere on E . By Theorems VIII.3.7 and VIII.3.8 it follows that the product $f\varphi$ is summable on E . The integral of this product is a functional expressed by the formula

$$F(f) = \int_E f\varphi d\mu \quad (31)$$

which is defined for all $f \in L$. By Theorem VIII.3.8, this functional possesses the distributive property and, besides, it is bounded; namely,

$$|F(f)| \leq C_0 \int_E |f| d\mu = C_0 \|f\| \quad (32)$$

Hence, by Theorem III.7.2, F is a linear functional.

It is obvious that all the functions equivalent to φ when substituted in formula (31) specify the same functional.

* Instead of "vrai sup" we also write "vrai max" or "essential sup" or "ess sup".

On the other hand, if two functions φ_1 and φ_2 , bounded almost everywhere and measurable, are not equivalent to each other it is easy to choose a function $f \in \mathcal{L}$ for which

$$\int_E f \varphi_1 d\mu \neq \int_E f \varphi_2 d\mu$$

which means that the linear functionals of type (31) specified by the functions φ_1 and φ_2 are different.

Let us show that the norm of linear functional (31) is given by the formula

$$\|F\| = \text{vrai sup } |\varphi(x)| \quad (33)$$

that is $\|F\| = C_0$.

From inequality (32) it follows that $\|F\| \leq C_0$. On the other hand, for an arbitrary $\varepsilon > 0$ there is a set $e \subset E$ with $\mu e > 0$ on which $|\varphi(x)| > C_0 - \varepsilon$. Passing, if necessary, to a subset $e' \subset e$ with $\mu e' < +\infty$ such that φ retains constant sign on e' , for instance, $\varphi(x) > 0$, we can take as f the characteristic function of the set e' . Then $\|f\| = \mu e'$ and

$$F(f) = \int_{e'} \varphi d\mu \geq (C_0 - \varepsilon) \mu e' = (C_0 - \varepsilon) \|f\|$$

It follows that $\|F\| \geq C_0 - \varepsilon$ and, since ε is arbitrary, $\|F\| \geq C_0$. Equality (33) has thus been proved.

It can be shown that formula (31) gives the general representation for a linear functional on $\mathcal{L}$; thus, every linear functional F on the space $\mathcal{L}$ can be expressed by formula (31) with an appropriately chosen measurable function φ bounded almost everywhere on E .*

§ 6. Geometrical Interpretation of Lebesgue's Integral in Euclidean Space

We shall elucidate the geometrical meaning of the Lebesgue integral for the special case of a Euclidean space. In the considerations below we shall simultaneously deal

* H. Steinhaus (1887-1972), a Polish mathematician, was the first to prove (in 1918) this theorem for the simplest case when $E = [0, 1] \subset \mathbb{R}_1$.

with Euclidean spaces of different dimensions and therefore we shall denote the Lebesgue measure in the space R_n by μ_n .

Let us introduce some geometrical notions. If a point $y = (\eta_1, \eta_2, \dots, \eta_n, \eta_{n+1}) \in R_{n+1}$ we shall call the point $x = (\eta_1, \eta_2, \dots, \eta_n)$, belonging to R_n , whose coordinates coincide with the first n coordinates of y , the *projection* of y into the space R_n . Furthermore, as a generalization of the notion of a curvilinear trapezoid we state the following definition.

Definition. Let $f(x) \geq 0$ be a function defined on a set $E \subset R_n$. Its *subgraph** on the set E is the collection Q of all the points $y = (\eta_1, \eta_2, \dots, \eta_n, \eta_{n+1}) \in R_{n+1}$ such that if x is the projection of the point y into R_n then

- (a) $x \in E$;
- (b) $0 \leq \eta_{n+1} \leq f(x)$.

In other words, for every $x \in E$ we construct a line segment $[0, f(x)]$ lying on the axis η_{n+1} , and Q is nothing but the collection of the points belonging to these segments.**

If $f(x) = c$ ($c \geq 0$ is a finite constant) on a set E the subgraph Q of the function f on E will be also called a *cylinder with base E and altitude c* . In such a case we shall write $Q = E \times [0, c]$ (cf. the notation for cells introduced in Sec. V.3).

Lemma VIII.6.1. A cylinder Q whose base is a measurable set $E \subset R_n$ and altitude is equal to c is a measurable set in the space R_{n+1} ; besides, $\mu_{n+1}Q = c\mu_n E$ for $c > 0$ and $\mu_{n+1}Q = 0$ for $c = 0$ ***.

Proof. Let us begin with a cylinder Q with altitude $c > 0$.

(a) Let the base of the cylinder Q be a nonempty open set $G \subset R_n$. By Theorem V.4.1, we have $G = \bigcup_k \Delta_k$ where Δ_k ($k = 1, 2, \dots$) are a countable number of disjoint non-empty n -dimensional cells. Then $Q = \bigcup_k (\Delta_k \times [0, c])$. Each of the sets on the right-hand side of the last relation

* The term "subgraph" was introduced by the Soviet mathematician I. P. Natanson (1906-1964).

** If $f(x) = +\infty$ then, as we agreed in Chapter V, the segment $[0, +\infty]$ corresponding to such a point x is understood as the half-open interval $[0, +\infty)$.

*** If $\mu_n E = +\infty$ and $c = 0$ the product $c\mu_n E$ does not make sense; but in this case as well we have $\mu_{n+1}Q = 0$.

is a parallelepiped with volume equal to $c\mu_n \Delta_h$. Therefore the cylinder Q is also measurable and, by the σ -additivity of measure,

$$\mu_{n+1}Q = \sum_h c\mu_n \Delta_h = c \sum \mu_n \Delta_h = c\mu_n G$$

(b) Now, let the base of the cylinder Q be a bounded closed set F . Let us embed F into a bounded open parallelepiped $\Delta \subset \mathbb{R}_n$. Then the set $G = \Delta \setminus F$ is also open. Consider the cylinders

$$Q_1 = G \times [0, c] \quad \text{and} \quad Q_2 = \Delta \times [0, c]$$

By what has been proved, the cylinders Q_1 and Q_2 are measurable, $\mu_{n+1}Q_1 = c\mu_n G$ and $\mu_{n+1}Q_2 = c\mu_n \Delta$. Finally, since $Q_1 \subset Q_2$ and $Q = Q_2 \setminus Q_1$ we conclude that the cylinder Q is measurable and

$$\mu_{n+1}Q = \mu_{n+1}Q_2 - \mu_{n+1}Q_1 = c(\mu_n \Delta - \mu_n G) = c\mu_n F$$

(c) Let the base of the cylinder Q be an arbitrary bounded measurable set E . Given an arbitrary $\varepsilon > 0$, we choose a closed set F and an open set G such that $F \subset E \subset G$ and $\mu_n G - \mu_n F < \varepsilon$ and then construct the cylinders Q' and Q'' with bases F and G , respectively, and altitude c . For these cylinders we have

$$Q' \subset Q \subset Q'', \quad \mu_{n+1}Q' = c\mu_n F, \quad \mu_{n+1}Q'' = c\mu_n G$$

Consequently

$$\mu_{n+1}Q'' - \mu_{n+1}Q' < c\varepsilon$$

Hence, the cylinder Q satisfies the relation $Q' \subset Q \subset Q''$ where Q' and Q'' are measurable sets for which the difference of their measures can be made arbitrarily small. According to the measurability criterion for a Euclidean space (see Sec. V.5), the cylinder Q is measurable as well, and therefore we can write

$$\mu_{n+1}Q' \leq \mu_{n+1}Q \leq \mu_{n+1}Q''$$

At the same time

$$\mu_{n+1}Q' \leq c \cdot \mu_n E \leq \mu_{n+1}Q''$$

the number $c\mu_n E$ being the only one lying between all the possible values of $\mu_{n+1}Q'$ and $\mu_{n+1}Q''$, that is $\mu_{n+1}Q = c \cdot \mu_n E$.

(d) Lastly, let the base of the cylinder Q be an arbitrary measurable set E . We can represent E in the form of the union of an increasing sequence of bounded measurable sets E_m ($E = \bigcup_{m=1}^{\infty} E_m$) and the cylinder Q in the form $Q = \bigcup_{m=1}^{\infty} Q_m$ where $Q_m = E_m \times [0, c]$. According to what has been proved, $\mu_{n+1}Q_m = c \cdot \mu_n E_m$, and the properties of measure imply

$$\mu_{n+1}Q = \lim \mu_{n+1}Q_m = c \cdot \lim \mu_n E_m = c \cdot \mu_n E$$

It now remains to consider the case when $c = 0$.

If $c = 0$ and $\mu_n E < +\infty$ we embed the cylinder Q into a cylinder Q' with the same base E and an arbitrarily small altitude $c' > 0$. Since the variation of c' makes it possible to obtain an arbitrarily small measure $\mu_{n+1}Q'$ the cylinder Q is measurable and $\mu_{n+1}Q = 0$ (see Theorem IV.4.2). If $c = 0$ and $\mu_n E = +\infty$ we represent E as a countable union of sets E_m with finite measure. Then $Q = \bigcup_{m=1}^{\infty} Q_m$ where Q_m ($m = 1, 2, \dots$) is a cylinder with base E_m and altitude $c = 0$. As was proved, $\mu_{n+1}Q_m = 0$, and therefore $\mu_{n+1}Q = 0$.

The proof of the lemma has been completed.

Theorem VIII.6.1. *If $f(x) \geq 0$ is a measurable function defined on a set $E \subset \mathbb{R}_n$ its subgraph Q on this set is a measurable set in $\mathbb{R}_{n+1}$ and*

$$\mu_{n+1}Q = \int_E f d\mu_n \quad (34)$$

Proof.* (a) Let us begin with the case when the function f is bounded on the set E and $\mu_n E < +\infty$. Consider an arbitrary partition τ of the set E into a finite number of disjoint measurable sets e_i ($i = 1, 2, \dots, p$). As usual, let us put

$$M_i = \sup_{x \in e_i} f(x), \quad m_i = \inf_{x \in e_i} f(x)$$

and form the Lebesgue-Darboux sums

$$S(\tau) = \sum_{i=1}^p M_i \mu_n e_i, \quad s(\tau) = \sum_{i=1}^p m_i \mu_n e_i$$

* This proof is valid both under the assumption that the function f is finite almost everywhere and without this assumption.

Let Q_i be the subgraph of the restriction of the function f to the set e_i , and Q'_i and Q''_i be, respectively, the cylinders with base e_i and altitudes m_i and M_i , and let $Q' = \bigcup_{i=1}^p Q'_i$, $Q'' = \bigcup_{i=1}^p Q''_i$. Then $Q'_i \subset Q_i \subset Q''_i$ for each i , and therefore $Q' \subset Q \subset Q''$, and we have

$$\left. \begin{aligned} \mu_{n+1}Q' &= \sum_{i=1}^p \mu_{n+1}Q'_i = \sum_{i=1}^p m_i \mu_n e_i = s(\tau) \\ \mu_{n+1}Q'' &= \sum_{i=1}^p \mu_{n+1}Q''_i = \sum_{i=1}^p M_i \mu_n e_i = S(\tau) \end{aligned} \right\} \quad (35)$$

The function f being integrable on the set E , an appropriate choice of the partition τ makes it possible to obtain an arbitrarily small value of the difference $S(\tau) - s(\tau)$. Consequently, by the measurability criterion, the set Q is measurable and

$$\mu_{n+1}Q' \leq \mu_{n+1}Q \leq \mu_{n+1}Q'' \quad (36)$$

On the other hand, we have

$$s(\tau) \leq \int_E f d\mu_n \leq S(\tau) \quad (37)$$

the integral $\int_E f d\mu_n$ being the only number lying between the lower and the upper Lebesgue-Darboux sums. Relations (35), (36) and (37) directly imply equality (34).

(b) Now let $\mu_n E = +\infty$ and f be bounded on E . In this case we represent E in the form of a countable union of an increasing sequence of sets E_m with finite measure ($E = \bigcup_{m=1}^{\infty} E_m$) and denote by Q_m the subgraph of the restriction of the function f to the set E_m . As has been proved,

$$\mu_{n+1}Q_m = \int_{E_m} f d\mu_n$$

and from formula (15) it follows that

$$\mu_{n+1}Q = \lim \mu_{n+1}Q_m = \lim \int_{E_m} f d\mu_n = \int_E f d\mu_n$$

(c) Finally, if f is unbounded we use the corresponding truncated functions f_m . The subgraphs Q_m of the truncated functions f_m form an increasing sequence, and $Q = \bigcup_{m=1}^{\infty} Q_m$. Therefore, using what has been proved in (a) and (b) and the property of the integrals of the truncated functions (see proposition (a) in Sec. VIII.4) we obtain

$$\mu_{n+1}Q = \lim \mu_{n+1}Q_m = \lim \int_E f_m d\mu_n = \int_E f d\mu_n$$

§ 7. Iterated Integrals

As in classical mathematical analysis, the Lebesgue integral over a set lying in a many-dimensional space can be computed by the reduction to an iterated (repeated) integral. We shall see that the ultimate result of the investigation presented below is in a certain sense more complete than that of the theory of the Riemann integral.

We shall begin with the problem of computing the measure of a set by integrating the measure of its sections. To simplify the notation we shall consider the case of a set lying in the two-dimensional space R_2 . At the end of this section we shall indicate how the result obtained for R_2 can be extended to the space R_n of an arbitrary dimension $n > 2$.

Let $E \subset R_2$ be an arbitrary set. For every fixed real number ξ_1 we shall denote by $E(\xi_1)$ the set of all ξ_2 's for which the point $x = (\xi_1, \xi_2) \in E$, i.e. each set $E(\xi_1)$ is the *section* of the set E by a straight line $\xi_1 = \text{const.}$ * We shall regard $E(\xi_1)$ as sets on the line and, when speaking of their measures, we shall mean the measures in the space R_1 .

Theorem VIII.7.1. *If a set $E \subset R_2$ is measurable, $\mu_2 E < +\infty$, the sections $E(\xi_1)$ are measurable and have finite measure for almost all ξ_1 , then the function $\mu_1 E(\xi_1)$ is measurable on R_1 and*

$$\mu_2 E = \int_{R_1} \mu_1 E(\xi_1) d\xi_1 \quad (38)$$

* More precisely, for given $\xi_1 = \text{const}$, the set $E(\xi_1)$ is the projection of the section made by the line $\xi_1 = \text{const}$ on the axis $O\xi_2$.

Proof. First of all we note that if E is an empty set formula (38) becomes evident. Our further argument reduces to the following stages.

(a) If we take, as E , a rectangle $\Delta = \langle a, b; c, d \rangle$ with finite edges then

$$E(\xi_1) = \begin{cases} \langle c, d \rangle & \text{for } a < \xi_1 < b \\ \emptyset & \text{for } \xi_1 < a \text{ and for } \xi_1 > b \end{cases}$$

Consequently,

$$\mu_1 E(\xi_1) = \begin{cases} d - c & \text{for } a < \xi_1 < b \\ 0 & \text{for } \xi_1 < a \text{ and for } \xi_1 > b, \end{cases}$$

and formula (38) obviously holds.

(b) Let G be an arbitrary open set in R_2 with $\mu_2 G < +\infty$. According to Theorem V.4.1, it is representable as a finite or countable union of disjoint two-dimensional cells Δ_k : $G = \bigcup_k \Delta_k$ ($\mu_2 \Delta_k < +\infty$). Then, for each ξ_1 , we have

$$G(\xi_1) = \bigcup_k \Delta_k(\xi_1), \quad \mu_1 G(\xi_1) = \sum_k \mu_1 \Delta_k(\xi_1)$$

In particular, all the sections $G(\xi_1)$ are measurable and the function $\mu_1 G(\xi_1)$ is measurable (see the corollary of Theorem VI.3.1). By the σ -additivity of measure and by formula (38) which has already been proved for rectangles we have

$$\mu_2 G = \sum_k \mu_2 \Delta_k = \sum_k \int_{R_1} \mu_1 \Delta_k(\xi_1) d\xi_1 \quad (39)$$

From (39) and from the corollary of Theorem VIII.4.2 (see proposition (c) in Sec. VIII.4) it follows immediately that $\mu_1 G(\xi_1) < +\infty$ for almost all ξ_1 ; besides, the signs of summation and integration on the right-hand side of (39) can be interchanged (cf. the same proposition (c)). We thus arrive at the equality

$$\mu_2 G = \int_{R_1} \mu_1 G(\xi_1) d\xi_1 \quad (40)$$

(c) If $E \subset R_2$ is a G_δ -set with $\mu_2 E < +\infty$ we can represent E in the form of the intersection of a decreasing sequence

of open sets G_m with finite measure:*

$$E = \bigcap_{m=1}^{\infty} G_m$$

Then $E(\xi_1) = \bigcap_{m=1}^{\infty} G_m(\xi_1)$ for any ξ_1 and the sets $G_m(\xi_1)$ also form a decreasing sequence. It follows that all the sections $E(\xi_1)$ are measurable and $\mu_1 E(\xi_1) = \lim \mu_1 G_m(\xi_1) < +\infty$ for almost all ξ_1 (since $\mu_1 G_1(\xi_1) < +\infty$ for almost all ξ_1). Now, by Theorem VIII.4.1, we conclude that

$$\mu_2 E = \lim \mu_2 G_m = \lim \int_{R_1} \mu_1 G_m(\xi_1) d\xi_1 = \int_{R_1} \mu_1 E(\xi_1) d\xi_1$$

(d) Let E be a measurable set in R_2 with $\mu_2 E = 0$. By Theorem V.5.6, there is a G_δ -set K such that $E \subset K$ and $\mu_2 K = 0$. According to the above,

$$\int_{R_1} \mu_1 K(\xi_1) d\xi_1 = \mu_2 K = 0$$

and, consequently, Theorem VIII.3.6 indicates that $\mu_1 K(\xi_1) = 0$ for almost all ξ_1 . But $E(\xi_1) \subset K(\xi_1)$ for each ξ_1 , and since the measure μ_1 is complete the sections $E(\xi_1)$ are measurable and $\mu_1 E(\xi_1) = 0$ for almost all ξ_1 . Whence, by Theorem VI.4.1, it follows that the function $\mu_1 E(\xi_1)$ is measurable and

$$\int_{R_1} \mu_1 E(\xi_1) d\xi_1 = 0 = \mu_2 E$$

(e) Finally, let E be an arbitrary measurable set in R_2 with $\mu_2 E < +\infty$. Choosing a G_δ -set K such that $E \subset K$ and $\mu_2(K \setminus E) = 0$ we obtain

$$E(\xi_1) = K(\xi_1) \setminus (K \setminus E)(\xi_1)$$

* If $E = \bigcap_{h=1}^{\infty} \Gamma_h$ where Γ_h are some open sets we put $G_m = \Gamma_1 \cap \Gamma_2 \cap \dots \cap \Gamma_m$. Then the sets G_m are also open; they constitute a decreasing sequence, and $E = \bigcap_{m=1}^{\infty} G_m$. Moreover, at the very beginning we can choose G_1 so that $\mu_2(G_1 \setminus E) < 1$ (see Corollary 1 of Theorem V.5.5), and then $\mu_2 G_1 < +\infty$.

and by what has been proved in (d), almost all the sections of the set $K \setminus E$ are of measure zero. Consequently, the sections $E(\xi_1)$ are measurable and $\mu_1 E(\xi_1) = \mu_1 K(\xi_1)$ for almost all ξ_1 , the function $\mu_1 E(\xi_1)$ is again measurable (Theorem VI.4.1) and

$$\int_{R_1} \mu_1 E(\xi_1) d\xi_1 = \int_{R_1} \mu_1 K(\xi_1) d\xi_1 = \mu_2 K = \mu_2 E$$

The proof of the theorem has been completed.

Remark. Formula (38) remains valid without the assumption that $\mu_2 E < +\infty$ but in this case we can no longer assert that $\mu_1 E(\xi_1) < +\infty$ for almost all sections. To obtain formula (38) for a set $E \subset R_2$ with $\mu_2 E = +\infty^*$ let us represent E as a union of an increasing sequence of measurable sets E_m having finite measure. Applying formula (38) to each set E_m we derive

$$\mu_2 E_m = \int_{R_1} \mu_1 E_m(\xi_1) d\xi_1 \quad (41)$$

It now becomes clear that $E(\xi_1) = \bigcup_{m=1}^{\infty} E_m(\xi_1)$ for every ξ_1 and consequently the sections $E(\xi_1)$ are measurable for almost all ξ_1 and $\mu_1 E_m(\xi_1)$ increases and tends to $\mu_1 E(\xi_1)$ as $m \rightarrow \infty$. Therefore Theorem VIII.4.2 (also see the remark made after this theorem) shows that it is allowable to pass to the limit under the sign of integration in equality (41) (for $m \rightarrow \infty$), which results in (38). As in Theorem VIII.7.1, the function $\mu_1 E(\xi_1)$ in this construction is measurable in R_1 .

Corollary. Let a nonnegative function f be defined on a measurable set $E \subset R_1$ and let its subgraph Q be a measurable set in R_2 . Then f is measurable.

Indeed, as was mentioned, the function $\mu_1 Q(\xi_1)$ is sure to be measurable on R_1 , and, in particular, on E . For any $\xi_1 \in E$ the section $Q(\xi_1)$ is the closed interval $[0, f(\xi_1)]$; therefore $\mu_1 Q(\xi_1) = f(\xi_1)$ and hence f is measurable.**

* The foregoing deduction of formula (38) can of course be carried out directly with some slight changes, without the assumption that $\mu_2 E < +\infty$.

** It follows from formula (38) that in the case $\mu_2 Q < +\infty$ the function f is summable.

Before proceeding to study iterated integrals we shall introduce some new notation and terminology. Let a function $f(\xi_1, \xi_2)^*$ be defined on a set $E \subset R_2$. If the section $E(\xi_1)$ is nonempty and measurable for some ξ_1 and if the integral of the function $f(\xi_1, \xi_2)$ regarded as a function of one variable ξ_2 over the set $E(\xi_1)$ makes sense we shall write this integral as

$$\int_{E(\xi_1)} f(\xi_1, \xi_2) d\xi_2 \quad (42)$$

If $E(\xi_1) = \emptyset$ for some ξ_1 then, although for such ξ_1 the function $f(\xi_1, \xi_2)$ is not defined for any ξ_2 , we shall agree that in this case as well integral (42) makes sense and is understood as being equal to zero.

Furthermore, let us agree that " $f(\xi_1, \xi_2)$ is measurable (or, respectively, summable) with respect to ξ_2 on the section $E(\xi_1)$ for almost all $\xi_1 \in R_1$ " means that for almost every $\xi_1 \in R_1$ either $E(\xi_1) \neq \emptyset$, $E(\xi_1)$ is measurable and $f(\xi_1, \xi_2)$ as function of ξ_2 is measurable (respectively, summable) on $E(\xi_1)$ or $E(\xi_1) = \emptyset$.

Theorem VIII.7.2 (G. Fubini).** *If $f(\xi_1, \xi_2)$ is summable on a set $E \subset R_2$ then it is summable as function of ξ_2 on $E(\xi_1)$ for almost all ξ_1 and*

$$\int_E f d\mu_2 = \int_{R_1} d\xi_1 \int_{E(\xi_1)} f(\xi_1, \xi_2) d\xi_2 \quad (43)$$

Here the inner integral is of course a measurable and even summable function of ξ_1 .

Proof. First, let $f(x) \geq 0$ on the set E and let Q be the subgraph of $f(x)$. By Theorem VIII.6.1,

$$\int_E f d\mu_2 = \mu_3 Q \quad (44)$$

and $\mu_3 Q < +\infty$. The set Q lies in the space R_3 , and the theorem analogous to Theorem VIII.7.1 holds for it. Con-

* In this notation we explicitly write both arguments of the function f which are the coordinates ξ_1 and ξ_2 of the variable point $x \in E \subset R_2$.

** G. Fubini (1879-1943), an Italian mathematician,

sequently, the plane section $Q(\xi_1)$ of the set Q^* is measurable for almost all ξ_1 and $\mu_2 Q(\xi_1) < +\infty$. We also have

$$\mu_3 Q = \int_{R_1} \mu_2 Q(\xi_1) d\xi_1 \quad (45)$$

Let us denote by A the set of all $\xi_1 \in R_1$ for which both sections $Q(\xi_1)$ and $E(\xi_1)$ are measurable and $\mu_2 Q(\xi_1) < +\infty$. From Theorem VIII.7.1 (and from the remark after this theorem) it follows that $\mu_1(R_1 \setminus A) = 0$. Moreover, if $E(\xi_1) \neq \emptyset$ then $Q(\xi_1)$ is the subgraph of the function $f(\xi_1, \xi_2)$ regarded as defined on $E(\xi_1)$ for fixed ξ_1 . Thus, if $\xi_1 \in A$ and $E(\xi_1) \neq \emptyset$ the corollary of Theorem VIII.7.1 implies that the function $f(\xi_1, \xi_2)$ is measurable as function of ξ_2 on the set $E(\xi_1)$; then it is also summable and

$$\mu_2 Q(\xi_1) = \int_{E(\xi_1)} f(\xi_1, \xi_2) d\xi_2 \quad (46)$$

In the case when $E(\xi_1) = \emptyset$ equality (46) is fulfilled automatically.

Comparing formulas (44), (45) and (46) we arrive at equality (43).

If $f(\xi_1, \xi_2)$ is an arbitrary function summable on the set E we consider the functions $f_+(\xi_1, \xi_2)$ and $f_-(\xi_1, \xi_2)$ (cf. formula (4)), write down formula (43) for each of them and then perform the termwise subtraction, which leads to the same formula for the original function $f(\xi_1, \xi_2)$.

The reduction of a double integral to an iterated integral by means of formula (43) can also be carried out without the requirement that f should be summable; it suffices to assume that the integral $\int_E f d\mu_2$ exists. In particular, there holds

the following theorem proved by the Italian mathematician L. Tonelli (1885-1946):

Theorem VIII.7.3. *Let $f(\xi_1, \xi_2) \geq 0$ be a measurable function on a set $E \subset R_2$.*

Then:

(i) *$f(\xi_1, \xi_2)$ is measurable as function of ξ_2 on the set $E(\xi_1)$ for almost all $\xi_1 \in R_1$;*

* $Q(\xi_1)$ is the set consisting of all the points $(\xi_2, \xi_3) \in R_2$ such that $(\xi_1, \xi_2, \xi_3) \in Q$.

(ii) formula (43) holds;

(iii) if the function $\varphi(\xi_1) = \int_{E(\xi_1)} f(\xi_1, \xi_2) d\xi_2$ is summable on R_1 the function $f(\xi_1, \xi_2)$ is summable on E .

To prove assertions (i) and (ii) it suffices to repeat the first part of the proof of Fubini's theorem without considering the section $Q(\xi_1)$ with a finite two-dimensional measure.* But in this case, instead of the summability of the function $f(\xi_1, \xi_2)$ with respect to ξ_2 for almost all ξ_1 , we can only guarantee that $f(\xi_1, \xi_2)$ is measurable and that the inner integral in (43) is a measurable function of ξ_1 . Assertion (iii) follows from formula (43):

$$\int_E f d\mu_2 = \int_{R_1} d\xi_1 \int_{E(\xi_1)} f(\xi_1, \xi_2) d\xi_2 = \int_{R_1} \varphi(\xi_1) d\xi_1 < +\infty$$

When computing the iterated integral in formula (43) we are in fact only interested in nonempty sections $E(\xi_1)$. The set of all ξ_1 for which $E(\xi_1) \neq \emptyset$ is called the *projection* of the set E on the first coordinate axis $O\xi_1$; we shall denote it $\text{proj}_1 E$. It is easy to see that the measurability of the set E in the space R_2 does not imply the measurability of $\text{proj}_1 E$ in the space R_1 . But if the set E is such that $\text{proj}_1 E$ is measurable the outer integral in formula (43) taken over R_1 can be replaced by the integral over $\text{proj}_1 E$ since

$$\int_{E(\xi_1)} f(\xi_1, \xi_2) d\xi_2 = 0 \quad \text{for } \xi_1 \notin \text{proj}_1 E$$

Therefore formula (43) can be rewritten as

$$\int_E f d\mu_2 = \int_{\text{proj}_1 E} d\xi_1 \int_{E(\xi_1)} f(\xi_1, \xi_2) d\xi_2$$

In particular, if the set E is a rectangle $\Delta = \langle a, b; c, d \rangle$ then

$$\int_{\Delta} f d\mu_2 = \int_a^b d\xi_1 \int_c^d f(\xi_1, \xi_2) d\xi_2$$

* Recall that Theorem VIII.7.4 on which the proof of Fubini's theorem is based also holds for sets of infinite measure.

We similarly obtain

$$\int_{\Delta} f d\mu_2 = \int_c^d d\xi_2 \int_a^b f(\xi_1, \xi_2) d\xi_1$$

It follows that if f is a summable function on a rectangle $\Delta = \langle a, b; c, d \rangle$ then

$$\int_a^b d\xi_1 \int_c^d f(\xi_1, \xi_2) d\xi_2 = \int_c^d d\xi_2 \int_a^b f(\xi_1, \xi_2) d\xi_1$$

Fubini's theorem indicates in which way the order of integration can be changed in more complicated cases when it is possible to represent the given iterated integral as the result of the computation of a double integral. In particular, for a nonnegative function it follows from Theorem VIII.7.3 that

if $f(\xi_1, \xi_2) \geq 0$ is a measurable function on a set $E \subset R_2$ then

$$\int_{R_1} d\xi_1 \int_{E(\xi_1)} f(\xi_1, \xi_2) d\xi_2 = \int_{R_1} d\xi_2 \int_{E(\xi_2)} f(\xi_1, \xi_2) d\xi_1 \quad (47)$$

To conclude this section we shall briefly discuss the generalization of Fubini's theorem to the case of the integral in a space of an arbitrary dimension exceeding two.

Let $f(\xi_1, \xi_2, \dots, \xi_n)$ be a summable function on a set $E \subset R_n$ with $\mu_n E < +\infty$. We divide the arguments of the function into two groups

$$\xi_1, \xi_2, \dots, \xi_k \quad (k < n) \text{ and } \xi_{k+1}, \dots, \xi_n$$

and, for every fixed collection of real numbers $\xi_1, \xi_2, \dots, \xi_k$, denote by

$$E(\xi_1, \xi_2, \dots, \xi_k)$$

the set of all the points $(\xi_{k+1}, \dots, \xi_n)$ of the space R_{n-k} for which

$$x = (\xi_1, \xi_2, \dots, \xi_k, \dots, \xi_n) \in E$$

Then it turns out that for almost all $(\xi_1, \dots, \xi_k) \in R_k$ either the function $f(\xi_1, \xi_2, \dots, \xi_k, \dots, \xi_n)$ is summable on $E(\xi_1, \xi_2, \dots, \xi_k)$ as a function of the variables

$\xi_{k+1}, \dots, \xi_n$ or $E(\xi_1, \xi_2, \dots, \xi_k) = \emptyset$. There also holds the equality

$$\int_E f d\mu_n = \int_{R_k} d\mu_k \int_{E(\xi_1, \xi_2, \dots, \xi_k)} f(\xi_1, \xi_2, \dots, \xi_k, \dots, \xi_n) d\mu_{n-k}^*$$

where, as before, we agree that in the case when $E(\xi_1, \dots, \xi_k) = \emptyset$ the inner integral is understood as equal to zero. In particular, for $k = 1$ we obtain

$$\int_E f d\mu_n = \int_{R_1} d\xi_1 \int_{E(\xi_1)} f(\xi_1, \xi_2, \dots, \xi_n) d\mu_{n-1}$$

Applying Fubini's theorem to the inner integral we derive the equality

$$\int_E f d\mu_n = \int_{R_1} d\xi_1 \int_{R_1} d\xi_2 \int_{E(\xi_1, \xi_2)} f(\xi_1, \xi_2, \dots, \xi_n) d\mu_{n-2}$$

Proceeding to transform the integrals in this way we finally obtain the representation of the integral over the set E in the form of an n -fold iterated integral involving n one-fold integrations. In particular, if the set E is a parallelepiped $\Delta = \langle a, b \rangle$ this representation assumes the simpler form

$$\int_{\Delta} f d\mu_n = \int_{a_1}^{b_1} d\xi_1 \int_{a_2}^{b_2} d\xi_2 \dots \int_{a_n}^{b_n} f(\xi_1, \xi_2, \dots, \xi_n) d\xi_n$$

§ 8. Product Measure

The theory presented in the foregoing section admits of a far-reaching generalization. Let X and Y be two measure spaces with measures μ and ν respectively, the σ -algebra of μ -measurable subsets of X being denoted by $\mathfrak{S}$ and the σ -algebra of ν -measurable subsets of Y by $\mathfrak{T}$. Both measures are supposed to be complete and σ -finite. Consider the set Z consisting of all ordered pairs (x, y) where $x \in X$ and $y \in Y$. The set Z is called the (*Cartesian*) *product* of the sets X and Y and is denoted $X \times Y$. Let us consider the collection

* Here the inner integration extends over "the collection of the variables $\xi_{k+1}, \dots, \xi_n$ ".

$\mathfrak{R}$ of all subsets of the set Z representable in the form

$$A \times B$$

where $A \in \mathfrak{S}$ and $B \in \mathfrak{T}$. It is readily seen that $\mathfrak{R}$ is a semiring. Now we define on $\mathfrak{R}$ the set function π by means of the formula

$$\pi(A \times B) = \mu A \cdot \nu B \quad (48)$$

on condition that if at least one of the factors on the right-hand side is equal to zero the product is also considered to be zero (even if the other factor is equal to $+\infty$). By analogy with Lemma V.2.1, changing appropriately the construction of the network partition we can easily prove that the function π is finitely additive.

On the other hand, the additivity of the function π can also be proved without referring to the construction of the partition by using the properties of the integral established above; this proof shows that π is not only finitely additive but also σ -additive.

Indeed, let

$$A \times B = \bigcup_n (A_n \times B_n)$$

where the sets on the right-hand side are disjoint ($A_n, A \in \mathfrak{S}$ and $B_n, B \in \mathfrak{T}$) and their number is finite or countable. For each n we now define on the set A the measurable function f_n given by the expression

$$f_n(x) = \begin{cases} \nu B_n & \text{for } x \in A_n \\ 0 & \text{for } x \in A \setminus A_n \end{cases}$$

Then it becomes clear that

$$\sum_n f_n(x) = \nu B$$

for all $x \in A$. Integrating this equality termwise (if there is a countable number of the factors we rely on Theorem VIII.4.2 and on the remark made after this theorem) we obtain

$$\sum_n \int_A f_n d\mu = \nu B \cdot \mu A = \pi(A \times B)$$

But we have

$$\int_A f_n d\mu = \nu B_n \cdot \mu A_n = \pi(A_n \times B_n)$$

and thus the σ -additivity of the function π has been proved. Hence, π is a measure. It is easily seen that it is σ -finite.

Now let us construct the standard extension of the measure π from the semiring $\mathfrak{R}$ to the corresponding σ -algebra which we shall denote by $\mathfrak{M}$. The measure thus constructed in $\mathbf{Z}$ will be denoted by $\mu \times \nu$ and referred to as the *product measure of the measures μ and ν* .*

In particular, let $\mathbf{X}$ and $\mathbf{Y}$ be the Euclidean spaces $\mathbf{R}_n$ and $\mathbf{R}_p$ with Lebesgue's measures μ_n and μ_p respectively. The product $\mathbf{R}_n \times \mathbf{R}_p$ can be interpreted as the Euclidean space $\mathbf{R}_{n+p}$. Let us show that *the product measure $\mu_n \times \mu_p$ in $\mathbf{R}_{n+p}$ coincides with the Lebesgue measure μ_{n+p}* .

Let $\mathfrak{R}$ be the semiring of the subsets of $\mathbf{R}_{n+p}$ of the form $A \times B$ where A and B are, respectively, measurable sets in $\mathbf{R}_n$ and in $\mathbf{R}_p$. It can readily be verified that such sets $A \times B$ are Lebesgue measurable in the space $\mathbf{R}_{n+p}$ **. Generalizing Theorem VIII.7.1 to spaces of arbitrary dimension we can easily prove that

$$\mu_{n+p}(A \times B) = \mu_n A \cdot \mu_p B$$

that is the function π specified by μ_n and μ_p with the aid of formula (48) coincides on $\mathfrak{R}$ with μ_{n+p} . As is known (see Theorem IV.4.3), the outer measure in $\mathbf{R}_{n+p}$ generated by the measure μ_{n+p} and the outer measure generated by the volumes of the cells in $\mathbf{R}_{n+p}$ coincide. Therefore the outer measure generated by the "intermediate" function π also coincides with them. It now follows that the standard extension of the measure π coincides with the Lebesgue measure μ_{n+p} .

Theorem VIII.6.1 elucidating the geometrical meaning of the integral in Euclidean spaces admits of the following generalization. Let $\mathbf{X}$ be an arbitrary measure space with

* We can similarly construct the measure in the product of any finite number of measure spaces. For an infinite collection of measures such a construction becomes more sophisticated.

** To this end it is sufficient to apply Corollary 1 of Theorem V.5.5.

a complete measure μ ; we put $Z = X \times R_1$ and $\omega = \mu \times \mu_1$ where μ_1 is the Lebesgue measure on the line. Then if $f(x) \geq 0$ is a measurable function on the set $E \subset X$ its subgraph

$$Q = \{(x, y) \in Z; \quad x \in E, \quad 0 \leq y \leq f(x)\}$$

is a measurable set in Z and

$$\omega Q = \int_E f d\mu$$

Let us come back to the general case considered at the beginning of the present section, that is let X and Y be arbitrary measure spaces with complete measures μ and ν , respectively. It can readily be checked that all the results established in Sec. VIII.7 for the space R_2 are extended to the space $Z = X \times Y$ with measure $\mu \times \nu$. In particular, formula (38) turns into

$$(\mu \times \nu) E = \int_X \nu E(x) d\mu$$

In the course of the proof of this formula one should use, instead of open sets, the sets representable as finite or countable unions of disjoint sets belonging to the semiring $\mathfrak{K}$. Accordingly, formula (43) is written as

$$\int_E f(x, y) d(\mu \times \nu) = \int_X d\mu \int_{E(x)} f(x, y) d\nu$$

(under the assumption that the integral on the left-hand side exists).

Square Summable Functions

§ 1. The Space L^2

In many applications of functional analysis the space L^2 of square summable functions plays a more important role than the space L . In its properties the former space turns out to be closer to the Euclidean space R_n than the latter.

As before, let X be a measure space with a measure μ defined on a σ -algebra of measurable sets. By a *square summable function on a set* $E \subset X$ we shall mean a measurable function f defined and finite almost everywhere on the set E such that

$$\int_E f^2 d\mu < +\infty$$

The collection of all square summable functions on a fixed measurable set $E \subset X$ will be denoted L^2 (or, more precisely, L^2_E). Thus, $f \in L^2_E$ (f is a function defined on E) if and only if f is measurable and $f^2 \in L_E$. Remember that in the set L we identified equivalent functions. In the set L^2 we shall also identify mutually equivalent functions. Therefore we can also regard every element of L^2 as being represented by a function having finite values everywhere on E .

We shall begin with proving that the set L^2 is a linear space. It is clear that if f belongs to L^2 the product cf also belongs to L^2 for any constant c . If f and g are two functions belonging to L^2 the obvious inequality

$$|f(x)g(x)| \leq f^2(x) + g^2(x)$$

shows that the product $fg \in L$. Then the equality

$$[f(x) + g(x)]^2 = f^2(x) + 2f(x)g(x) + g^2(x)$$

implies that the function $(f + g)^2$ is summable and thus $f + g$ enters into L^2 . Consequently, the set L^2 is a linear

(function) space; moreover, the product of any two functions belonging to L^2 is summable (but, at the same time, it may not belong to L^2). As in the space L , the zero element θ of the space L^2 is represented by any measurable function equivalent to 0.

We shall derive some inequalities which are an integral analogue of Cauchy's inequality (see Sec. II.1). Let $f, g \in L^2$. Consider the auxiliary function φ of the real variable λ determined by the formula

$$\varphi(\lambda) = \int_E (\lambda f + g)^2 d\mu = \lambda^2 \int_E f^2 d\mu + 2\lambda \int_E fg d\mu + \int_E g^2 d\mu$$

where λ assumes all the values from $-\infty$ to $+\infty$. Since $\lambda f + g \in L^2$ for any λ the function φ takes finite values. Furthermore, $\varphi(\lambda) \geq 0$ for all λ since the integral of a non-negative function is nonnegative. Since $\varphi(\lambda)$ is a quadratic trinomial assuming nonnegative values for all λ , its discriminant must not exceed 0 (see Sec. II.1), that is

$$\left(\int_E fg d\mu \right)^2 \leq \int_E f^2 d\mu \int_E g^2 d\mu \quad (1)$$

This relation is known as *Bunyakovsky's* inequality*.

Now, by analogy with Sec. II.1, we extract the square roots of both members of Bunyakovsky's inequality, multiply them by 2 and add to them expression

$$\int_E f^2 d\mu + \int_E g^2 d\mu$$

This results in

$$\int_E (f + g)^2 d\mu \leq \left(\sqrt{\int_E f^2 d\mu} + \sqrt{\int_E g^2 d\mu} \right)^2$$

that is

$$\sqrt{\int_E (f + g)^2 d\mu} \leq \sqrt{\int_E f^2 d\mu} + \sqrt{\int_E g^2 d\mu} \quad (2)$$

* V. Ya. Bunyakovsky (1804-1889), a noted Russian mathematician.

Let us introduce the norm in the space L^2 by putting

$$\|f\| = \sqrt{\int_E f^2 d\mu}^*$$

Then inequality (2) assumes the form $\|f + g\| \leq \|f\| + \|g\|$, i.e. turns into the triangle inequality. The fulfilment of all the other axioms of norm in the space L^2 is obvious, and thus L^2 equipped with such a norm is a normed space.

If $X = R_n$, μ is the Lebesgue measure and $\mu E > 0$ (the case $\mu E = 0$ is obviously of no interest) it is always possible to construct a function summable on E but having nonsummable square. Hence, L is not contained in L^2 . If $\mu E = +\infty$ the space L^2 is not a part of L either. For instance, if E is the interval $(1, +\infty) \subset R_1$ the function $\frac{1}{x}$ is nonsummable on E but belongs to L^2_E . In this connection the following theorem is of special importance.

Theorem IX.1.1. *If $\mu E < +\infty$ then $L^2 \subset L$ and the inequality*

$$\|f\|_L \leq \sqrt{\mu E} \cdot \|f\|_{L^2} \quad (3)$$

holds for any $f \in L^2$ (here the subscripts L and L^2 in the notation of the norms indicate the spaces in which the corresponding norms are taken).**

Proof. If $\mu E < +\infty$ any function identically equal to a constant on the set E belongs to L^2 . Therefore the product of any function f belonging to L^2 by any constant is summable on E . In particular, the function f is itself summable. Therefore, by Bunyakovsky's inequality,

$$\left(\int_E |f| d\mu \right)^2 \leq \int_E f^2 d\mu \int_E 1^2 d\mu$$

which is equivalent to inequality (3).

* Here, of course, we mean the arithmetic root.

** Our discussion shows that in the general case a square summable function f is not necessarily summable. In the case $L^2 \subset L$ both f and f^2 are summable.

The convergence $f_k \rightarrow f$ with respect to the norm in the space $\mathbf{L}^2$ means that

$$\int_E |f_k - f|^2 d\mu \rightarrow 0$$

This type of convergence is spoken of as the *convergence in the mean of order two* or the *convergence in the mean square*. For the convergence there holds a theorem analogous to Theorem VIII.5.1 proved for the convergence in the mean of order one:

Theorem IX.1.2. *If f_k converges in the mean square to f as $k \rightarrow \infty$ (i.e. converges in $\mathbf{L}^2$) then $f_k \Rightarrow f$.*

The proof of this theorem is almost the same as in the case of Theorem VIII.5.1. If we suppose that the sequence $\{f_k\}$ is not convergent in measure to f then there are some $\varepsilon > 0$ and $\delta > 0$ such that

$$\mu E(|f_k(x) - f(x)| \geq \varepsilon) \geq \delta$$

for an infinite number of values $k_1, k_2, \dots, k_i, \dots$ of the index k .

It then follows that

$$\int_E |f_{k_i} - f|^2 d\mu \geq \delta \varepsilon^2$$

and we thus arrive at a contradiction to the convergence in norm of $\{f_k\}$ to f .

Theorem IX.1.3. *If $\mu E < +\infty$ and $f_k \rightarrow f$ with respect to the norm of $\mathbf{L}^2$ then $f_k \rightarrow f$ with respect to the norm of $\mathbf{L}$, that is the convergence in the mean square implies the convergence in the mean of order one to the same limit function.*

Proof. For $\mu E < +\infty$ we have $\mathbf{L}^2 \subset \mathbf{L}$, and inequality (3) shows that

$$\|f_k - f\|_{\mathbf{L}} \leq \sqrt{\mu E} \|f_k - f\|_{\mathbf{L}^2}$$

whence follows the assertion of the theorem.

Theorem IX.1.4. *$\mathbf{L}^2$ is a Banach space.*

To prove the theorem we should only establish the com-

pleteness of the space L^2 , which is achieved in the same manner as in the case of the space L (see Theorem VIII.5.2).*

If $X = R_n$ and μ is the Lebesgue measure we can follow the scheme used for the space L to prove the separability of the space L^2 . Namely, that very countable set H which was constructed in the course of the proof of the separability of L (Sec. VIII.5) turns out to be everywhere dense in the space L^2 as well. To prove this one should make the following changes in the proof applied to L : inequality (27) in Chapter VIII should be replaced by the inequality

$$\int_{E \setminus E_p} f^2 d\mu < \frac{\varepsilon^2}{4} \text{ and the set } e \subset E \text{ should be chosen so that}$$

$$\int_{E_p \setminus e} f^2 d\mu < \varepsilon^2/16. \text{ Then inequality (28) of Chapter VIII is replaced by}$$

$$\int_{E_p} |f - g|^2 d\mu < \frac{\varepsilon^2}{16}$$

The continuous function φ is chosen so that

$$\mu E_p [g(x) \neq \varphi(x)] < \frac{\varepsilon^2}{64M^2}$$

Inequality (29) of Chapter VIII then goes into

$$\int_{E_p} |g - \varphi|^2 d\mu < \frac{\varepsilon^2}{16}$$

Finally, the polynomial P must be constructed so that

$$|\varphi(x) - P(x)| < \frac{\varepsilon}{4\sqrt{\mu\Delta_p^*}} \text{ throughout } \Delta_p^*. \text{ Then, instead of inequality (30), Chapter VIII, we obtain}$$

$$\int_{E_p} |\varphi - \psi|^2 d\mu \leq \frac{\varepsilon^2}{16}$$

The final estimation of the norm $\|f - \psi\|$ (in the space L^2) can be carried out as follows. We have

$$\|f - \psi\|^2 = \int_{E_p} |f - \psi|^2 d\mu + \int_{E \setminus E_p} f^2 d\mu$$

* The space L^2 was investigated by mathematicians earlier than the space L . The completeness of L^2 was proved at the beginning of the 20th century by E. S. Fischer (1875-1954), a German mathematician.

By inequality (2),

$$\begin{aligned} \sqrt{\int_{E_p} |f - \psi|^2 d\mu} &\leq \sqrt{\int_{E_p} |f - g|^2 d\mu} + \\ &+ \sqrt{\int_{E_p} |g - \varphi|^2 d\mu} + \sqrt{\int_{E_p} |\varphi - \psi|^2 d\mu} < \frac{3}{4} \varepsilon \end{aligned}$$

and therefore $\|f - \psi\|^2 < \frac{13}{16} \varepsilon^2$ which implies $\|f - \psi\| < \varepsilon$.

§ 2. Scalar Product

The space $\mathbf{L}^2$ resembles $\mathbf{R}_n$ not only because the distance in $\mathbf{L}^2$ is determined by the formula analogous to the formula for the distance between two points in $\mathbf{R}_n$ (the only difference is that the role of the sum is played by the integral) but also because it is possible to introduce the notion of the scalar product of two elements of $\mathbf{L}^2$, which makes the analogy still closer.

Definition. The *scalar product* of two functions f and g belonging to the space $\mathbf{L}^2$ (denoted as (f, g)) is defined as the integral of the product of the functions:

$$(f, g) = \int_E fg d\mu \quad (4)$$

As was proved in the foregoing section, the scalar product (f, g) has a finite value for any two functions $f, g \in \mathbf{L}^2$.

Recall that in $\mathbf{R}_n$ the scalar product of two vectors $x = (\xi_1, \xi_2, \dots, \xi_n)$ and $y = (\eta_1, \eta_2, \dots, \eta_n)$ is given by the formula

$$(x, y) = \sum_{i=1}^n \xi_i \eta_i \quad (5)$$

Therefore it becomes clear that formula (4) can be regarded as an analogue of formula (5) obtained by the replacement of a finite number of the coordinates of the vectors by an infinite number of the values of the functions with the substitution of the integral sign for the summation sign. By the way, the space $\mathbf{L}^2$ can be interpreted not only as an analogue of the space $\mathbf{R}_n$: if the space $\mathbf{L}_E^2$ is constructed on a

set E consisting of n points and the measure μ in E is defined so that the measure of every single-element set is equal to 1, the space L_E^2 turns into R_n and the general formula (4) goes into formula (5); this means that R_n is a special case of L^2 .

When considering the scalar product (f, g) we sometimes simply call the elements f and g *factors entering into the product*. It appears evident that if one or both factors entering into the scalar product are replaced by equivalent functions this does not affect the value of the scalar product.

The following properties of the scalar product are an immediate consequence of its definition:

- (a) $(f, g) = (g, f)$;
- (b) $(f + g, h) = (f, h) + (g, h)$;
- (c) $(cf, g) = c(f, g)$ where c is any real number*;
- (d) $(f, f) \geq 0$ for any function $f \in L^2$ and $(f, f) = 0$ if and only if $f = \theta$.

Properties (b) and (c) are a special case of the more general formula

$$\left(\sum_{i=1}^k \alpha_i f_i, \sum_{j=1}^l \beta_j g_j \right) = \sum_{i=1}^k \sum_{j=1}^l \alpha_i \beta_j (f_i, g_j)$$

(which expresses the *distributive property* of the scalar product). The Bunyakovsky inequality can be written as

$$(f, g)^2 \leq \|f\|^2 \|g\|^2$$

or in the form

$$|(f, g)| \leq \|f\| \|g\| \quad (6)$$

It is clear that the norm and the scalar product in L^2 are connected by the relation

$$\|f\| = \sqrt{(f, f)} \quad (7)$$

Every Banach space in which with its any two elements f and g there is associated a real number called their scalar product (denoted (f, g)) so that properties (a)-(d) hold and the norm of any element is expressed in terms of the scalar product by formula (7) is referred to as a *real Hilbert space***. Thus, L^2 is a real Hilbert (function) space.

* Properties (b) and (c) also apply to the second factors.

** After the great German mathematician D. Hilbert (1862-1943).

Let us show that the scalar product is a continuous function of its arguments, that is, if $f_k \rightarrow f$ and $g_k \rightarrow g$ in norm in L^2 then $(f_k, g_k) \rightarrow (f, g)$.

Indeed, by virtue of the distributivity of the scalar product, we have the identity

$$(f_k, g_k) - (f, g) = (f_k - f, g_k) + (f, g_k - g)$$

The norms of the elements g_k being uniformly bounded (since g_k form a convergent sequence; see Sec. III.6), i.e. $\|g_k\| \leq C$ for all k , it follows from inequality (6) that

$$|(f_k - f, g_k)| \leq \|f_k - f\| \|g_k\| \leq C \|f_k - f\| \rightarrow 0$$

and

$$|(f, g_k - g)| \leq \|f\| \|g_k - g\| \rightarrow 0$$

whence $(f_k, g_k) \rightarrow (f, g)$.

The continuity of the scalar product makes it possible to prove that the distributive property continues to hold for an infinite number of summands (on infinite series see

Sec. III.6). Namely, if $f = \sum_{i=1}^{\infty} f_i$ and g is an arbitrary element of L^2 we have

$$(f, g) = \lim_{k \rightarrow \infty} \left(\sum_{i=1}^k f_i, g \right) = \lim_{k \rightarrow \infty} \sum_{i=1}^k (f_i, g) = \sum_{i=1}^{\infty} (f_i, g) \quad (8)$$

Definition. Two functions f and g belonging to L_E^2 are said to be *orthogonal* on the set E (written $f \perp g$) if $(f, g) = 0$.

From the properties of the scalar product it follows immediately that

- (i) the zero element θ is orthogonal to any function $f \in L^2$;
- (ii) $f \perp f$ if and only if $f = \theta$;
- (iii) if $f = \sum_i c_i f_i$ (the sum involves a finite or countable number of summands) and $g \perp f_i$ for all i then $f \perp g$ (in the case of a countable number of summands this follows from formula (8));

(iv) if a set A is everywhere dense in L^2 and f is orthogonal to every function belonging to A then $f = \theta$. Indeed, by the definition of an everywhere dense set, $f = \lim f_k$ where $f_k \in A$. The continuity of the scalar product then implies $(f, f) = \lim (f, f_k)$ whence $f = \theta$.

For the space L^2 there holds the *generalized Pythagoras' theorem*: if $f = \sum_i f_i$ (the sum contains a finite or countable number of summands) and all the elements f_i are pairwise orthogonal then

$$\|f\|^2 = \sum_i \|f_i\|^2$$

Indeed, by the distributivity of the scalar product,

$$\|f\|^2 = (f, f) = \left(\sum_i f_i, \sum_j f_j\right) = \sum_i \sum_j (f_i, f_j)$$

But $(f_i, f_j) = 0$ for $i \neq j$ and therefore

$$\|f\|^2 = \sum_i (f_i, f_i) = \sum_i \|f_i\|^2$$

Theorem IX.2.1. For a series $\sum_{i=1}^{\infty} f_i$ consisting of pairwise orthogonal elements to be convergent (in norm) it is necessary and sufficient that the number series $\sum_{i=1}^{\infty} \|f_i\|^2$ converge.

Proof. If the series $\sum_{i=1}^{\infty} f_i$ is convergent then, by Pythagoras' theorem, the series $\sum_{i=1}^{\infty} \|f_i\|^2$ is also convergent.

Conversely, let the number series $\sum_{i=1}^{\infty} \|f_i\|^2$ be convergent.

Putting $s_k = \sum_{i=1}^k f_i$ we can write, for $l > k$,

$$s_l - s_k = \sum_{i=k+1}^l f_i$$

whence, by Pythagoras' theorem,

$$\|s_l - s_k\|^2 = \sum_{i=k+1}^l \|f_i\|^2$$

It follows that $\|s_l - s_k\| \rightarrow 0$ for $k, l \rightarrow \infty$, i.e. $\{s_k\}$ is a Cauchy sequence. By the completeness of the space L^2 ,

there exists $\lim s_k$, that is the series $\sum_{i=1}^{\infty} f_i$ is convergent.

§ 3. Orthogonal Series

In this section we shall discuss the problem of expanding functions belonging to L^2 into orthogonal series, that is into series composed of pairwise mutually orthogonal functions. When speaking of convergence of series we shall always mean convergence in the mean of order two (convergence in the mean square), that is the convergence in norm in the space L^2 .

Definition. A system of functions $\varphi_1, \varphi_2, \dots, \varphi_i, \dots$ (finite or countable) belonging to L_E^2 is called *orthonormal* on the set E if $\varphi_i \perp \varphi_j$ for $i \neq j$ and $\|\varphi_i\| = 1$ for all i .

In what follows we shall only deal with infinite orthonormal systems, and this fact will not be stipulated repeatedly. The set E we shall deal with will be regarded as fixed.

If ψ_i are pairwise orthogonal functions on the set E and are different from θ (such a system of functions is simply called *orthogonal*) the (normalized) functions $\varphi_i = \frac{\psi_i}{\|\psi_i\|}$ form an orthonormal system.

An important example of an orthonormal system on the interval $[-\pi, \pi] \subset \mathbf{R}_1$ is the well-known trigonometric system of functions

$$\frac{1}{\sqrt{2\pi}}, \quad \frac{1}{\sqrt{\pi}} \cos x, \quad \frac{1}{\sqrt{\pi}} \sin x, \dots, \quad \frac{1}{\sqrt{\pi}} \cos kx, \\ \frac{1}{\sqrt{\pi}} \sin kx, \dots \quad (9)$$

For the interval $[-1, 1]$ we can indicate the orthonormal system of polynomials

$$p_k(x) = \frac{1}{2^k k!} \sqrt{\frac{2k+1}{2}} \frac{d^k}{dx^k} [(x^2 - 1)^k] \quad (10)$$

* These polynomials differ from Legendre's polynomials P_k , widely applied in mathematics, only in the factors $\sqrt{\frac{2k+1}{2}}$. On Legendre's polynomials see, for instance, B. M. Budak, S. V. Fomin, *Multiple Integrals, Field Theory and Series*, Moscow, Mir Publishers, 1973 (English edition).

For the square $[-\pi, \pi; -\pi, \pi] \subset R_2$ an example of an orthonormal system is the one formed of the functions

$$\frac{1}{\pi} \cos k\xi_1 \sin l\xi_2 \quad (k, l = 1, 2, \dots)$$

Orthogonal series possess the following important property: if $\{\varphi_i\}$ is an orthonormal system of functions belonging to L^2 and f is a function representable as $f = \sum_{i=1}^{\infty} \alpha_i \varphi_i$ (as usual, the sum of the series is understood in the sense of the convergence in norm) the coefficients α_i are determined uniquely, namely, they are given by the formulas $\alpha_i = (f, \varphi_i)$ ($i = 1, 2, \dots$).

Indeed, by the distributivity of the scalar product,

$$(f, \varphi_j) = \sum_{i=1}^{\infty} \alpha_i (\varphi_i, \varphi_j) = \alpha_j \|\varphi_j\|^2 = \alpha_j$$

for any $j = 1, 2, \dots$

Definition. Let there be given an orthonormal system of functions $\{\varphi_i\}$ belonging to L^2 . For an arbitrary function $f \in L^2$ the scalar products (f, φ_i) ($i = 1, 2, \dots$) are termed the *Fourier coefficients of the function f with respect to the given orthonormal system* and the series $\sum_{i=1}^{\infty} \alpha_i \varphi_i$ where $\alpha_i = (f, \varphi_i)$ ($i = 1, 2, \dots$) is called the *Fourier series of the function f with respect to that system*.

Theorem IX.3.1. *The Fourier series of any function f belonging to L^2 converges in norm. For its sum to coincide (in the space L^2) with f it is necessary and sufficient that the equality*

$$\|f\|^2 = \sum_{i=1}^{\infty} \alpha_i^2 \tag{11}$$

be fulfilled where α_i are the Fourier coefficients of the function f .

Proof. For an arbitrary natural number k , let us put

$$g = f - \sum_{i=1}^k \alpha_i \varphi_i$$

Then, for any $j = 1, 2, \dots, k$, we have

$$(g, \varphi_j) = (f, \varphi_j) - \sum_{i=1}^k \alpha_i (\varphi_i, \varphi_j) = \alpha_j - \alpha_j \|\varphi_j\|^2 = 0$$

that is g is orthogonal to all φ_j ($j = 1, 2, \dots, k$). Now, the representation of f in the form $f = g + \sum_{i=1}^k \alpha_i \varphi_i$ and Pythagoras' theorem imply that

$$\|f\|^2 = \|g\|^2 + \sum_{i=1}^k \alpha_i^2$$

Consequently, $\sum_{i=1}^k \alpha_i^2 \leq \|f\|^2$ for any k , whence follow the

convergence of the series $\sum_{i=1}^{\infty} \alpha_i^2$ and the inequality

$$\sum_{i=1}^{\infty} \alpha_i^2 \leq \|f\|^2 \quad (12)$$

Therefore, by Theorem IX.2.1, the series $\sum_{i=1}^{\infty} \alpha_i \varphi_i$ is convergent in norm.

If $f = \sum_{i=1}^{\infty} \alpha_i \varphi_i$ equality (11) follows from Pythagoras' theorem. Conversely, let equality (11) be fulfilled. Putting $h = f - \sum_{i=1}^{\infty} \alpha_i \varphi_i$ we readily prove, by analogy with what was shown for the function g , that $h \perp \varphi_i$ for all i , and therefore, applying Pythagoras' theorem again, we obtain

$$\|f\|^2 = \|h\|^2 + \sum_{i=1}^{\infty} \alpha_i^2$$

Since, by the hypothesis, equality (11) is fulfilled we have $h = 0$, i.e. $f = \sum_{i=1}^{\infty} \alpha_i \varphi_i$. The theorem has been proved.

We have incidentally shown that for any function $f \in L^2$ relation (12) holds; it is called *Bessel's inequality*.* Equality

* After F. W. Bessel (1784-1846), a German astronomer.

(11) is a special case of (12); it is referred to as *Parseval's* relation*. An orthonormal system satisfying Parseval's relation (11) for every $f \in L^2$ is said to be *closed* (this is another meaning of the word "closed", not to be confused with its meaning in Sec. III.3).

The following important theorem is a mere consequence of the results already established.

Theorem IX.3.2 (F. Riesz and E. S. Fisher). *Let $\{\varphi_i\}$ be an orthonormal system of functions belonging to L^2 and α_i*

be real numbers such that $\sum_{i=1}^{\infty} \alpha_i^2 < +\infty$. Then

(i) *the series $\sum_{i=1}^{\infty} \alpha_i \varphi_i$ converges in the mean square to a function f belonging to L^2 ;*

(ii) *α_i 's are the Fourier coefficients of the function f ;*

(iii) *Parseval's relation holds for the function f .*

Indeed, the convergence of the series $\sum_{i=1}^{\infty} \alpha_i \varphi_i$ follows from Theorem IX.2.1, and, by the uniqueness of the Fourier expansion, it is the Fourier series of its sum. Parseval's relation is implied by Theorem IX.3.1.

Definition. An orthonormal system of functions $\{\varphi_i\} \subset L^2$ is said to be *complete* if there is no function belonging to L^2 , different from θ and orthogonal to all φ_i .

It is obvious that if at least one function is removed from a complete system the remaining system is no longer complete.

If $\{\varphi_i\}$ is a complete system then there cannot exist two different (i.e. not equivalent to each other) functions belonging to L^2 which have one and the same Fourier series. Indeed, if two functions f and g have one and the same system of Fourier's coefficients their difference is orthogonal to all φ_i and hence $f - g = \theta$.

Theorem IX.3.3. *For an orthonormal system $\{\varphi_i\}$ to possess the property that the Fourier series with respect to this system of any function f belonging to L^2 is convergent in the mean of order two to that function it is necessary and sufficient that the system $\{\varphi_i\}$ be complete.*

* M. A. Parseval (?-1836), a French mathematician.

Proof. Let the system $\{\varphi_i\}$ be complete. Taking $f \in L^2$ and constructing the function h as was done in the course of the proof of Theorem IX.3.1 we see that, by virtue of the completeness of the system $\{\varphi_i\}$, the element h coincides with θ . But this exactly means that f is the sum of its Fourier series.

If the system $\{\varphi_i\}$ is not complete there is a function f belonging to L^2 , different from θ and orthogonal to all φ_i . All the Fourier coefficients of this function are zero and, consequently, the sum of its Fourier series is equal to θ and thus is different from f .

Let us show that trigonometric system (9) is complete in the space L^2 constructed on the interval $[-\pi, \pi]$. To this end we first prove that the set T of all *trigonometric polynomials*

$$P(x) = a_0 + \sum_{k=1}^m (a_k \cos kx + b_k \sin kx)$$

is everywhere dense in $L^2_{[-\pi, \pi]}$.

Let us take an arbitrary function $f \in L^2_{[-\pi, \pi]}$. Given any $\varepsilon > 0$, we choose a continuous function g defined on the interval $[-\pi, \pi]$ such that $\|f - g\| < \frac{\varepsilon}{3}$. The existence of such a function follows from the proof of the separability of the space L^2_E (for $E \subset \mathbb{R}_n$) presented in Sec. IX.1*. Let M be an upper bound of g : $|g(x)| < M$. Next we choose a positive number δ such that the inequalities $\delta < \frac{\varepsilon^2}{36M^2}$ and $\delta < 2\pi$ hold simultaneously.

Let us take the continuous function h defined by the relations

$$h(x) = g(x) \text{ for } -\pi \leq x \leq \pi - \delta$$

$$h(\pi) = h(-\pi) = g(-\pi)$$

and extend $h(x)$ to the interval $(\pi - \delta, \pi)$ by means of linear interpolation. Then we also have $|h(x)| < M$, and

$$\|g - h\| = \sqrt{\int_{\pi-\delta}^{\pi} |g(x) - h(x)|^2 dx} < \sqrt{4M^2\delta} < \frac{\varepsilon}{3}$$

* Since the set E on which the space L^2 in question is constructed is a finite interval the algebraic polynomials with rational coefficients form an everywhere dense set in L^2 (see the footnote on p. 216).

By Weierstrass' trigonometric approximation theorem*, there is a trigonometric polynomial P for which

$$|h(x) - P(x)| < \frac{\varepsilon}{3\sqrt{2\pi}}$$

throughout the interval $[-\pi, \pi]$. Therefore it can easily be shown that $\|h - P\| < \frac{\varepsilon}{3}$. Hence, $\|f - P\| < \varepsilon$.

Consequently, the set T is everywhere dense in $L^2_{[-\pi, \pi]}$.

Now it is easy to establish the completeness of trigonometric system (9). If a function $f \in L^2_{[-\pi, \pi]}$ is orthogonal to all the functions of system (9) it is orthogonal to every trigonometric polynomial as well and therefore, according to proposition (iv) in Sec. IX.2, we have $f = 0$.

The proof of the completeness of the system of polynomials (10) in the space L^2 constructed on the interval $[-1, 1]$ is still simpler.

Generally, it can be proved that in any separable space L^2_E (with $\mu E > 0$) there is a complete orthonormal system**.

Theorem IX.3.4. *If f is a function representable by its Fourier series, that is $f = \sum_{i=1}^{\infty} \alpha_i \varphi_i$ in the sense of the convergence in norm, then for any measurable set $e \subset E$ with $\mu e < +\infty$ the relation*

$$\int_e f d\mu = \sum_{i=1}^{\infty} \alpha_i \int_e \varphi_i d\mu \quad (13)$$

holds.

This means that it is allowable to integrate the expansion of a function into Fourier's series not only when the series fails to converge uniformly but even when it is not pointwise convergent.

Proof. Let g be the characteristic function of the set e . Then $g \in L^2_E$, and the distributivity of scalar product shows that $(f, g) = \sum_{i=1}^{\infty} \alpha_i (\varphi_i, g)$, which means that equality (13) holds.

* E.g. see B. M. Budak, S. V. Fomin, *Multiple Integrals, Field Theory and Series*, Moscow, Mir Publishers, 1973 (English edition).

** E.g. see B. Z. Vulikh, *An Introduction to Functional Analysis*, Moscow, "Nauka", 1967 (Russian edition), Theorem 6.9.1.

The existence of a complete orthonormal system in a separable space L^2 makes it possible to establish an interesting connection between such a space and the space l^2 (see Sec. III.1). First of all we note that the space l^2 can be turned into a normed one if we equip it with the norm

$$\|x\| = \sqrt{\sum_{i=1}^{\infty} \xi_i^2}$$

where $x = \{\xi_i\}$ is its arbitrary element. The verification of the fulfilment of the axioms of norm is quite simple in this case. In particular, the triangle inequality for this norm turns into Cauchy's inequality for infinite sequences (see formula (1), Chapter III). The algebraic operations in the space l^2 are a special case of algebraic operations in a general linear space of sequences (see Sec. III.6).

Let $\{\varphi_i\}$ ($i = 1, 2, \dots$) be a complete orthonormal system of functions in L^2 . Taking an arbitrary function $f \in L^2$ we can associate with it the sequence of its Fourier coefficients $\alpha_i = (f, \varphi_i)$. According to Bessel's inequality (12), the number sequence $\{\alpha_i\}$ is an element of the space l^2 , that is $\alpha = \{\alpha_i\} \in l^2$. From Theorems IX.3.3 and IX.3.4 it then follows directly that $\|f\| = \|\alpha\|$. We have thus constructed a norm-preserving mapping of the space L^2 into the space l^2 .

From the Riesz-Fisher Theorem (see Theorem IX.3.2) it follows immediately that this is a mapping of the space L^2 onto l^2 , that is every sequence α belonging to l^2 is the image of a function $f \in L^2$. Moreover, by the completeness of the system $\{\varphi_i\}$, the mapping thus constructed is one-to-one, that is every sequence $\alpha = \{\alpha_i\} \in l^2$ specifies a unique function (to within the equivalence of functions in L^2) for which α_i 's are its Fourier coefficients.

This one-to-one mapping of the space L^2 onto l^2 possesses the following properties:

(a) if $f, g \in L^2$ and $\alpha = \{\alpha_i\}$, $\beta = \{\beta_i\}$ are respectively, the sequences of the Fourier coefficients of f and g then to the sum $f + g$ there corresponds the sum $\alpha + \beta = \{\alpha_i + \beta_i\}$;

(b) if $f \in L^2$, α is the sequence corresponding to f and c is a real number then to the function cf there corresponds the sequence $c\alpha = \{c\alpha_i\}$.

When there is a one-to-one mapping of a normed space onto another normed space preserving the norm and possessing properties (a) and (b) we say that these spaces are *isomorphic* and *isometric*.

We have thus shown that *any separable space L^2 and the space l^2 are isomorphic and isometric*.

In the space l^2 we can introduce a scalar product by putting $(\alpha, \beta) = \sum_{i=1}^{\infty} \alpha_i \beta_i$ (the scalar product specified by this formula possesses properties (a)-(d) indicated in Sec. IX.2). It can also be readily checked that for any $f, g \in L^2$ and the sequences $\alpha, \beta \in l^2$ corresponding to them there holds the equality

$$(f, g) = (\alpha, \beta)$$

§ 4. Linear Functionals on L^2

Let us fix the factor g in the scalar product (f, g) and make the f factor run throughout the space L^2 . Then (f, g) can be regarded as a functional $F(f)$ with the argument $f \in L^2$. The properties of the scalar product indicate that this functional is distributive, and the Bunyakovsky inequality shows that

$$|F(f)| \leq \|f\| \cdot \|g\|$$

Consequently, F is a bounded functional and, by Theorem III.7.2, it is a linear functional on L^2 with $\|F\| \leq \|g\|$.

We shall prove that there holds in fact the more precise relation $\|F\| = \|g\|$. Indeed, for $f = g$ we have

$$F(g) = (g, g) = \|g\|^2 = \|g\| \cdot \|g\|$$

whence it follows that $\|F\| \geq \|g\|$; together with the foregoing estimation this leads to the desired equality.

It turns out that the converse is also true, that is every linear functional F in the space L^2 is representable as the scalar product (f, g) with an appropriately chosen function g . The proof of this fact is preceded by the following lemma.

Lemma IX.4.1. *If F is a linear functional on L^2 not identically equal to zero and H is the collection of all $f \in L^2$ such that $F(f) = 0$, there exists a function $h \neq 0$ belonging to L^2 which is orthogonal to all $f \in H$.**

Proof. The functional F not being identically zero, there is a function $\varphi \in L^2$ not belonging to H . Let us put

$$d = \inf_{f \in H} \|f - \varphi\|$$

and prove that the infimum is attained on a function $f_0 \in H$, that is there exists $f_0 \in H$ such that $\|f_0 - \varphi\| = d$. To this end, we choose, for every natural number n , $f_n \in H$ such that $\|f_n - \varphi\| < d_n$ where $d_n = d + \frac{1}{n}$. We shall show that $\{f_n\}$ is a Cauchy sequence.

It is verified directly that for any $u, v \in L^2$ we have

$$\|u + v\|^2 + \|u - v\|^2 = 2(\|u\|^2 + \|v\|^2)$$

Substituting $u = f_n - \varphi$ and $v = f_m - \varphi$ into this relation we obtain

$$\begin{aligned} \|(f_n + f_m) - 2\varphi\|^2 + \|f_n - f_m\|^2 &= \\ &= 2(\|f_n - \varphi\|^2 + \|f_m - \varphi\|^2) \end{aligned} \quad (14)$$

But

$$(f_n + f_m) - 2\varphi = 2 \left[\frac{1}{2} (f_n + f_m) - \varphi \right]$$

and $\frac{1}{2}(f_n + f_m) \in H$; consequently $\|(f_n + f_m) - 2\varphi\| \geq 2d$. Now it follows from (14) that

$$\|f_n - f_m\|^2 \leq 2(d_n^2 + d_m^2) - 4d^2 \xrightarrow{n, m \rightarrow \infty} 0$$

The completeness of L^2 therefore implies that there exists $f_0 = \lim f_n$, and by the continuity of F , we have $F(f_0) = 0$, that is $f_0 \in H$. For any n we can write

$$d \leq \|f_n - \varphi\| < d_n$$

Now, using the continuity of the norm and passing to the limit in the last inequality we immediately obtain $\|f_0 - \varphi\| = d$.

* This lemma continues to hold when H is an arbitrary closed linear subspace of L^2 different from L^2 .

Let us put $h = f_0 - \varphi$ and prove that $h \perp f$ for any function $f \in H$. It should be noted that $h \neq \theta$ since $\varphi \notin H$. If $f = \theta$ the relation $h \perp f$ is fulfilled automatically. Therefore we shall assume that $f \neq \theta$. For any real λ the function $f_0 - \lambda f$ belongs to H , and consequently

$$\|h - \lambda f\|^2 = \|(f_0 - \lambda f) - \varphi\|^2 \geq d^2 \quad (15)$$

Since $\|h\| = d$ the direct computation shows that

$$\|h - \lambda f\|^2 = d^2 - 2\lambda (h, f) + \lambda^2 \|f\|^2$$

whence, by (15), it follows that

$$\lambda^2 \|f\|^2 - 2\lambda (h, f) \geq 0 \quad (16)$$

Let us take $\lambda = \frac{(h, f)}{\|f\|^2}$. Then from (16) we obtain

$$-\frac{(h, f)^2}{\|f\|^2} \geq 0$$

which is only possible when $(h, f) = 0$, that is we have in fact $h \perp f$. The lemma is proved.

Now we are able to prove the general representation theorem for a linear functional in L^2 .

Theorem IX.4.1. (R. M. Fréchet-F. Riesz). *The general representation of a linear functional defined on the space L^2 (i.e. L^2_E) is given by the formula*

$$F(f) = \int_E fg \, d\mu \quad (17)$$

where g is also an element of L^2 . The norm of F is specified by the formula

$$\|F\| = \sqrt{\int_E g^2 \, d\mu} \quad (18)$$

and the functional F determines the function g uniquely (to within the equivalence of functions in L^2).

Proof. As has been shown, if $g \in L^2$ formula (17) specifies a linear functional F on L^2 for which equality (18) is fulfilled.

Now let F be an arbitrary linear functional on L^2 . If $F(f) \equiv 0$ the functional F is representable by formula (17)

with $g(x) \equiv 0$. Therefore we shall suppose that the functional F is not identically equal to zero. Let us apply the lemma proved above and take the function $h \in \mathcal{L}^2$ whose existence has been established in this lemma. If, as in the proof of the lemma, we put

$$H = \mathcal{L}^2 [F(f) = 0]$$

then

$$F(f)h - F(h)f \in H$$

for any $f \in \mathcal{L}^2$ and therefore the scalar product of $F(f)h - F(h)f$ by h turns into zero:

$$(F(f)h - F(h)f, h) = 0,$$

On computing the left-hand side we find

$$F(f) \|h\|^2 - F(h)(f, h) = 0$$

whence

$$F(f) = \frac{F(h)}{\|h\|^2} (f, h)$$

If we now put

$$g = \frac{F(h)}{\|h\|^2} h$$

this immediately leads to

$$F(f) = (f, g)$$

that is the possibility of the representation of the functional F by means of formula (17) has been proved.

It only remains to verify the uniqueness of the function g . Let $g_1 \neq g_2$. Consider the functionals

$$F_1(f) = (f, g_1)$$

and

$$F_2(f) = (f, g_2)$$

For $f_0 = g_1 - g_2$ we have

$$F_1(f_0) - F_2(f_0) = (g_1 - g_2, g_1 - g_2) = \|g_1 - g_2\|^2 > 0$$

and thus $F_1(f_0) \neq F_2(f_0)$; consequently the functionals F_1 and F_2 do not coincide.

The Space L^p § 1. The Definition of the Space L^p Basic Inequalities

In this chapter we shall deal with spaces which are a direct generalization of the space of square summable functions. As in the previous chapter, X will denote a measure space with measure μ and E a fixed measurable set belonging to X . Let us choose a number $p > 1$ and denote by L^p (or, more precisely, L^p_E) the collection of all functions f measurable and finite almost everywhere on the set E such that

$$\int_E |f|^p d\mu < +\infty^*$$

As usual, we identify in the set L^p mutually equivalent functions. In what follows we assume that every element of L^p is represented by a function taking on finite values.

It can readily be shown that L^p is a linear space. Indeed, if $f \in L^p$ then obviously $cf \in L^p$ for any constant c . Furthermore, let $f, g \in L^p$. Putting

$$E_1 = E[|f(x)| \leq |g(x)|] \text{ and } E_2 = E \setminus E_1$$

we can write

$$|f(x) + g(x)|^p \leq (|f(x)| + |g(x)|)^p \leq 2^p |g(x)|^p$$

for any $x \in E_1$ and therefore

$$\int_{E_1} |f+g|^p d\mu \leq 2^p \int_{E_1} |g|^p d\mu < +\infty$$

* The measurability of the function $|f|^p$ is an obvious consequence of the measurability of f .

Similarly,

$$\int_{E_2} |f + g|^p d\mu \leq 2^p \int_{E_2} |f|^p d\mu < +\infty$$

Hence,

$$\int_E |f + g|^p d\mu < +\infty$$

which means that $f + g \in L^p$.

In the study of the space L^p an essential role is played by another space L^q of the same type (on the same set E) whose exponent q is connected with p by the relation

$$\frac{1}{p} + \frac{1}{q} = 1 \quad (1)$$

Such exponents p and q are termed (mutually) *conjugate*. From (1) it follows that $q = \frac{p}{p-1} > 1$. Note that if $p = 2$ then also $q = 2$. But if $p \neq 2$ then $q \neq p$.

Let us derive some inequalities generalizing inequalities (1) and (2) of the foregoing chapter. We shall begin with proving the inequality

$$a^{\frac{1}{p}} b^{\frac{1}{q}} \leq \frac{a}{p} + \frac{b}{q} \quad (2)$$

valid for any real numbers $a, b \geq 0$ (here q is the exponent conjugate to p).

If at least one of the numbers a and b is equal to 0 inequality (2) becomes obvious and therefore we shall assume that $a, b > 0$. Let us put $m = \frac{1}{p}$ (then $0 < m < 1$) and consider the function

$$\varphi(t) = t^m - mt \quad (t > 0)$$

We have $\varphi'(t) = m(t^{m-1} - 1)$, and consequently, $\varphi'(t) > 0$ for $t < 1$ and $\varphi'(t) < 0$ for $t > 1$. This means that the function φ increases in the interval $(0, 1]$, decreases in the interval $[1, +\infty)$ and achieves the greatest value at the point $t = 1$. Consequently, $\varphi(t) \leq \varphi(1)$ for all $t > 0$; hence, $t^m - mt \leq 1 - m$, i.e. $t^m - 1 \leq m(t - 1)$.

Let us put $t = \frac{a}{b}$. Then the last inequality turns into

$$\frac{a^m}{b^m} - 1 \leq m \left(\frac{a}{b} - 1 \right)$$

whence, on multiplying termwise by b , we derive

$$a^m b^{1-m} \leq ma + (1 - m) b$$

But $m = \frac{1}{p}$ and $1 - m = \frac{1}{q}$, and thus we obtain inequality (2).

Let f and g be two arbitrary functions belonging, respectively, to the spaces L^p and L^q . Suppose first that both integrals

$$I_1 = \int_E |f|^p d\mu \quad \text{and} \quad I_2 = \int_E |g|^q d\mu$$

exceed zero. In inequality (2) we put

$$a = \frac{1}{I_1} |f(x)|^p \quad \text{and} \quad b = \frac{1}{I_2} |g(x)|^q$$

Then it follows from (2) that

$$|f(x) g(x)| \leq I_1^{\frac{1}{p}} I_2^{\frac{1}{q}} \left\{ \frac{|f(x)|^p}{p I_1} + \frac{|g(x)|^q}{q I_2} \right\}$$

Since the right-hand member is a summable function on the set E Theorem VIII.3.7 shows that the product fg is also summable and

$$\left| \int_E fg d\mu \right| \leq I_1^{\frac{1}{p}} I_2^{\frac{1}{q}} \left(\frac{1}{p} + \frac{1}{q} \right) = I_1^{\frac{1}{p}} I_2^{\frac{1}{q}}$$

This inequality continues to hold when I_1 or I_2 (or both) is equal to 0 since in this case f or g (or both) is equivalent to zero. Thus, the product fg of any two functions $f \in L^p$ and $g \in L^q$ is summable and there holds the relation

$$\left| \int_E fg d\mu \right| \leq \left(\int_E |f|^p d\mu \right)^{\frac{1}{p}} \left(\int_E |g|^q d\mu \right)^{\frac{1}{q}} \quad (3)$$

known as Hölder's* inequality. Bunyakovsky's inequality (see inequality (1) in Chapter IX) is a special case of Hölder's inequality.

Now let f and g be two arbitrary functions belonging to L^p . We shall put

$$I_1 = \int_E |f|^p d\mu, \quad I_2 = \int_E |g|^p d\mu, \quad I = \int_E |f+g|^p d\mu$$

and prove that

$$I^{\frac{1}{p}} \leq I_1^{\frac{1}{p}} + I_2^{\frac{1}{p}} \quad (4)$$

In the case when $I = 0$ inequality (4) is obvious and therefore we shall suppose that $I \neq 0$. Let q be the conjugate exponent to p . Since $|f+g|^p$ is summable the function

$|f+g|^{\frac{p}{q}}$ belongs to L^q . Note that $\frac{p}{q} = p - 1$.

Furthermore,

$$|f(x) + g(x)|^p \leq (|f(x)| + |g(x)|) |f(x) + g(x)|^{p-1} \quad (5)$$

Now applying Hölder's inequality to the product $|f| |f+g|^{p-1}$ and taking into account that $(p-1)q = p$ we obtain

$$\int_E |f| |f+g|^{p-1} d\mu \leq I_1^{\frac{1}{p}} I^{\frac{1}{q}}$$

Similarly,

$$\int_E |g| |f+g|^{p-1} d\mu \leq I_2^{\frac{1}{p}} I^{\frac{1}{q}}$$

It now follows from (5) that

$$I \leq \left(I_1^{\frac{1}{p}} + I_2^{\frac{1}{p}} \right) I^{\frac{1}{q}}$$

which leads to inequality (4). Thus, given any two functions $f, g \in L^p$, there holds the inequality

$$\left(\int_E |f+g|^p d\mu \right)^{\frac{1}{p}} \leq \left(\int_E |f|^p d\mu \right)^{\frac{1}{p}} + \left(\int_E |g|^p d\mu \right)^{\frac{1}{p}}$$

* O. Hölder (1859-1937), a German mathematician.

known as *Minkowski's* inequality*. The triangle inequality which holds for L^2 is a special case of Minkowski's inequality.

To define the norm in L^p we put

$$\|f\| = \left(\int_E |f|^p d\mu \right)^{\frac{1}{p}}$$

Minkowski's inequality then turns into the triangle inequality for the space L^p . The verification of all the other axioms of norm in the space L^p is quite obvious and consequently L^p equipped with this norm is a normed space.

Note that in the definition of the space L^p we can assume that $p \geq 1$. In the case when $p = 1$, the space L^p turns into the space L of summable functions which was considered in Sec. VIII.5.

Theorem X.1.1. *If $\mu E < +\infty$ inequality $s > r \geq 1$ implies that $L^s \subset L^r$, and for any function $f \in L^s$ the relation*

$$\|f\|_{L^r} \leq (\mu E)^{\frac{s-r}{rs}} \|f\|_{L^s} \quad (6)$$

*holds.***

Proof. Let us put $p = \frac{s}{r}$. Then the conjugate exponent is $q = \frac{s}{s-r}$. If $f \in L^s$ the function $|f|^r$ belongs to L^p . At the same time, since $\mu E < +\infty$ any constant (regarded as a function defined on the set E) belongs to L^q . Therefore the product of the function $|f|^r$ by any constant is summable on E . In particular, the function $|f|^r$ is itself summable, and thus $f \in L^r$. Moreover, Hölder's inequality shows that

$$\begin{aligned} \int_E |f|^r d\mu &\leq \left(\int_E |f|^s d\mu \right)^{\frac{1}{p}} \left(\int_E 1^q d\mu \right)^{\frac{1}{q}} = \\ &= (\mu E)^{\frac{s-r}{s}} \left(\int_E |f|^s d\mu \right)^{\frac{r}{s}} \end{aligned}$$

* H. Minkowski (1864-1909), a German mathematician and physicist.

** Here the indices in the notation of the norms indicate the spaces in which the norms are taken (cf. Chapter IX).

On raising both members of the last inequality to the $\frac{1}{r}$ th power we arrive at inequality (6).

The convergence $f_k \rightarrow f$ with respect to the norm of the space L^p means that

$$\int_E |f_k - f|^p d\mu \rightarrow 0$$

In this case we speak of *the convergence in the mean of order p* or *the mean convergence of the p th order*.

The theorems below are a direct generalization of the theorems proved in Sec. IX.1 for the space L^2 . We shall limit ourselves to the formulation of the theorems since their proofs are completely analogous to those of the corresponding theorems in Sec. IX.1.

Theorem X.1.2. *If f_k converges in norm to f in the space L^p then $f_k \Rightarrow f$.*

Theorem X.1.3. *If $\mu E < +\infty$ and $s > r \geq 1$ then the convergence $f_k \rightarrow f$ in norm in L^s implies that $f_k \rightarrow f$ with respect to the norm of L^r , that is the convergence in the mean of a certain order implies the convergence in the mean of any lower order to the same limit function*.*

Theorem X.1.4. L^p is a Banach space for any $p \geq 1$. If $E \supset X = \mathbb{R}_n$ and μ is the Lebesgue measure the space L^p_E is separable.

§ 2. Mean Value of a Function

If the space L^r is constructed on a set E with $0 < \mu E < +\infty$ then for each function $f \in L^r$ there exists *the mean value of order r of its modulus* which we shall denote by $M_r(f)$:

$$M_r(f) = \left\{ \frac{1}{\mu E} \int_E |f|^r d\mu \right\}^{1/r} \quad (7)$$

* We have only defined the convergence in the mean with exponents $p \geq 1$. The notion of the mean convergence of the p th order can be defined similarly when $0 < p < 1$. On the other hand, this type of convergence no longer admits of the interpretation as the convergence in norm since Minkowski's inequality fails to hold for the exponents $p < 1$.

It is advisable to include the case $\mu E = +\infty$ and therefore, instead of the ordinary mean value determined by formula (7), we shall consider the so-called weighted mean value. Namely, let us choose a bounded function α , summable on the set E , such that $\alpha(x) > 0$ for all $x \in E$. Then for every $r \geq 1$ and any $f \in L^r$ the product $\alpha|f|^r$ is summable, and we can define the *weighted mean* of the modulus of the function f by putting

$$M_r(f) = \left\{ \frac{\int_E \alpha |f|^r d\mu}{\int_E \alpha d\mu} \right\}^{\frac{1}{r}} \quad (8)$$

The function α entering into this definition is referred to as a *weight function*. The mean value determined by formula (7) is obviously a special case of (8) and is obtained from (8) when $\alpha(x) \equiv 1$ and $\mu E < +\infty$.

For the sake of brevity let us introduce a modified weight function

$$\beta(x) = \frac{\alpha(x)}{\int_E \alpha d\mu}$$

Then $\int_E \beta d\mu = 1$ and formula (8) assumes the shortened form

$$M_r(f) = \left\{ \int_E \beta |f|^r d\mu \right\}^{\frac{1}{r}}$$

Thus, the weighted means computed with the aid of the functions α and β coincide.

The mean value $M_r(f)$ can also be considered for any function f measurable on the set E but if $f \notin L^r$ it is not excluded that $M_r(f) = +\infty$.

Theorem X.2.1. *If $1 \leq r < s$ then $M_r(f) \leq M_s(f)$ for any function f measurable on the set E .*

Proof. If $M_s(f) = +\infty$ the assertion of the theorem holds automatically. Let us suppose that $M_s(f) < +\infty$.

Putting $p = \frac{s}{r}$ and taking the exponent q conjugate to p

we can represent the product $\beta |f|^r$ in the form

$$\beta |f|^r = (\beta |f|^s)^{\frac{1}{p}} \beta^{\frac{1}{q}}$$

Here the first factor on the right-hand side belongs to L^p while the other factor belongs to L^q . By Hölder's inequality,

$$\begin{aligned} (M_r(f))^r &= \int_E \beta |f|^r d\mu \leq \\ &\leq \left(\int_E \beta |f|^s d\mu \right)^{\frac{1}{p}} \left(\int_E \beta d\mu \right)^{\frac{1}{q}} = (M_s(f))^{\frac{s}{p}} \end{aligned}$$

whence it follows immediately that $M_r(f) \leq M_s(f)$, which is what we set out to prove.

Let us extend the definition of the essential supremum stated in Sec. VIII.5 to the case when the function f is not bounded almost everywhere. To this end we put $\text{vrai sup } |f(x)| = +\infty$ for such a function f .

Theorem X.2.2. *For any function f measurable on the set E the limiting relation*

$$M_r(f) \xrightarrow{r \rightarrow +\infty} \text{vrai sup } |f(x)| \quad (9)$$

holds.

Proof. The existence of $\lim M_r(f)$ follows from the foregoing theorem. Let $C = \text{vrai sup } |f(x)|$. If C is finite it is clear that

$$M_r(f) \leq C \left(\int_E \beta d\mu \right)^{\frac{1}{r}} = C$$

for any r . The same inequality $M_r(f) \leq C$ becomes obvious if $C = +\infty$.

Let us take an arbitrary number $C_1 < C$. Then there is a subset $E_1 \subset E$ with $\mu E_1 > 0$ such that $|f(x)| > C_1$ on it. Now we put

$$\tau = \int_{E_1} \beta d\mu$$

Since $\beta(x) > 0$ everywhere we have $\tau > 0$. Therefore it is clear that

$$M_r(f) \geq \left\{ \int_{E_1} \beta |f|^r d\mu \right\}^{\frac{1}{r}} \geq C_1 \tau^{\frac{1}{r}}$$

On passing to the limit for $r \rightarrow \infty$ we see that

$$\lim M_r(f) \geq C_1$$

The arbitrariness of the choice of the number C_1 now shows that

$$\lim M_r(f) \geq C$$

which completes the proof of relation (9).

§ 3. An Inequality for Iterated Integrals

In this section we shall prove a theorem concerning iterated integrals. The proof will be elaborated for integrals over sets on the line with finite measure but it should be noted that the ultimate result established in this theorem can readily be extended to integrals over arbitrary measurable sets in Euclidean spaces of arbitrary dimension*.

Let E_1 and E_2 be some measurable sets in $\mathbf{R}_1$ having a finite measure. According to the general definition of the (Cartesian) product of sets, by $E = E_1 \times E_2$ is meant the set consisting of all the points $(\xi_1, \xi_2) \in \mathbf{R}_2$ for which $\xi_1 \in E_1$ and $\xi_2 \in E_2$. As was mentioned in Sec. VIII.8, E is a measurable set in $\mathbf{R}_2$, and, therefore, by Theorem VIII.7.1, $\mu E = \mu E_1 \cdot \mu E_2$. Using these notions we now state the following theorem.

Theorem X.3.1. *Let $f(\xi_1, \xi_2)$ be a measurable nonnegative function defined on E . Suppose that*

(a) *$f(\xi_1, \xi_2)$ as function of one argument ξ_2 , for a fixed value of ξ_1 , is summable on E_2 for almost all $\xi_1 \in E_1$;*

* And also to integrals in the product spaces with a complete measure (see Sec. VIII.8).

(b) there is some $r > 1$ such that the integral of f^r over E_1 satisfies the condition

$$\int_{E_1} f^r(\xi_1, \xi_2) d\xi_1 < +\infty$$

for almost all $\xi_2 \in E_2$.

Then

$$\left\{ \int_{E_1} \left(\int_{E_2} f(\xi_1, \xi_2) d\xi_2 \right)^r d\xi_1 \right\}^{\frac{1}{r}} \leq \int_{E_2} \left(\int_{E_1} f^r(\xi_1, \xi_2) d\xi_1 \right)^{\frac{1}{r}} d\xi_2 \quad (10)$$

If, for a given r , condition (b) fails to hold, inequality (10) is nevertheless (automatically) fulfilled since in this case we should consider the right member of (10) as equal to $+\infty$. If condition (a) is violated both members of inequality (10) are equal to $+\infty$.

Inequality (10) is an integral analogue of the inequality

$$\left\| \sum_i f_i \right\|_{L^r} \leq \sum_i \|f_i\|_{L^r}$$

involving finite or countable sums which, in its turn, is a generalization of the triangle inequality for the norm in the space L^r .

Proof. Let us begin with the truncated functions f_m corresponding to the function f . The r th powers of the moduli of the truncated functions are summable on E , and therefore (see Theorem VIII.7.2) the integrals

$$\int_{E_2} f_m(\xi_1, \xi_2) d\xi_2 \text{ and } \int_{E_1} f_m^r(\xi_1, \xi_2) d\xi_1$$

are summable functions of the variables ξ_1 and ξ_2 respectively.

Let us put

$$\varphi_m(\xi_1) = \int_{E_2} f_m(\xi_1, \xi_2) d\xi_2$$

Then $\varphi_m(\xi_1) \leq m\mu E_2$, and consequently $\varphi_m \in L^r_{E_1}$. We also have

$$\int_{E_1} \varphi_m^r(\xi_1) d\xi_1 = \int_{E_1} \left(\varphi_m^{r-1}(\xi_1) \int_{E_2} f_m(\xi_1, \xi_2) d\xi_2 \right) d\xi_1 \quad (11)$$

The function $\varphi_m^{r-1}(\xi_1) f_m(\xi_1, \xi_2)$ is bounded and measurable on E and hence it is summable on this set. Therefore Fubini's theorem implies

$$\begin{aligned} \int_E \varphi_m^{r-1}(\xi_1) f_m(\xi_1, \xi_2) d\mu_2 &= \int_{E_1} \varphi_m^{r-1}(\xi_1) d\xi_1 \int_{E_2} f_m(\xi_1, \xi_2) d\xi_2 = \\ &= \int_{E_2} d\xi_2 \int_{E_1} \varphi_m^{r-1}(\xi_1) f_m(\xi_1, \xi_2) d\xi_1 \end{aligned} \quad (12)$$

We apply Hölder's inequality with $p = \frac{r}{r-1}$ to the last integral over the set E_1 (then the corresponding conjugate exponent is $q = r$), which results in

$$\begin{aligned} \int_{E_1} \varphi_m^{r-1}(\xi_1) f_m(\xi_1, \xi_2) d\xi_1 &\leq \\ &\leq \left(\int_{E_1} \varphi_m^r(\xi_1) d\xi_1 \right)^{\frac{1}{p}} \left(\int_{E_1} f_m^r(\xi_1, \xi_2) d\xi_1 \right)^{\frac{1}{r}} \end{aligned} \quad (13)$$

From (11), (12) and (13) it now follows

$$\int_{E_1} \varphi_m^r(\xi_1) d\xi_1 \leq \left(\int_{E_1} \varphi_m^r(\xi_1) d\xi_1 \right)^{\frac{1}{p}} \int_{E_2} \left(\int_{E_1} f_m^r(\xi_1, \xi_2) d\xi_1 \right)^{\frac{1}{r}} d\xi_2$$

that is

$$\left(\int_{E_1} \varphi_m^r(\xi_1) d\xi_1 \right)^{\frac{1}{r}} \leq \int_{E_2} \left(\int_{E_1} f_m^r(\xi_1, \xi_2) d\xi_1 \right)^{\frac{1}{r}} d\xi_2 \quad (14)$$

We have thus proved inequality (10) for the truncated functions.

Let us come back to the function f and put

$$\varphi(\xi_1) = \int_{E_2} f(\xi_1, \xi_2) d\xi_2$$

By condition (a), $\varphi(\xi_1) < +\infty$ for almost all $\xi_1 \in E_1$. The property of truncated functions established in proposition (a), Sec. VIII.4 indicates that

$$\varphi_m(\xi_1) \rightarrow \varphi(\xi_1) \quad \text{for almost all } \xi_1 \in E_1$$

Furthermore, $\varphi_{m+1}(\xi_1) \geq \varphi_m(\xi_1)$ ($m = 1, 2, \dots$). By Levi's theorem (Sec. VIII.4.2),

$$\int_{E_1} \varphi_m^r(\xi_1) d\xi_1 \xrightarrow{m \rightarrow \infty} \int_{E_1} \varphi^r(\xi_1) d\xi_1$$

On the other hand, the same theorem gives us

$$\int_{E_1} f_m^r(\xi_1, \xi_2) d\xi_1 \xrightarrow{m \rightarrow \infty} \int_{E_1} f^r(\xi_1, \xi_2) d\xi_1$$

for almost all $\xi_2 \in E_2$. Applying Levi's theorem once again we obtain

$$\int_{E_2} \left(\int_{E_1} f_m^r(\xi_1, \xi_2) d\xi_1 \right)^{\frac{1}{r}} d\xi_2 \xrightarrow{m \rightarrow \infty} \int_{E_2} \left(\int_{E_1} f^r(\xi_1, \xi_2) d\xi_1 \right)^{\frac{1}{r}} d\xi_2$$

It now becomes clear that the desired inequality (10) is readily obtained by means of the passage to the limit in (14).

§ 4. Linear Functionals on L^p

Let $p > 1$ and let q be the conjugate exponent to p . If g is a function belonging to L^q its product by any function $f \in L^p$ is summable and, consequently, the integral of this product specifies a functional on L^p expressed by the formula

$$F(f) = \int_E fg d\mu \quad (15)$$

The distributive property of this functional is quite evident and, according to Hölder's inequality, we have

$$|F(f)| \leq \|f\|_{L^p} \|g\|_{L^q}$$

Thus, F is a bounded functional and hence, by Theorem III.7.2, it is a linear functional on L^p , its norm satisfying the inequality $\|F\| \leq \|g\|_{L^q}$. We shall show that there holds in fact the exact equality

$$\|F\| = \|g\|_{L^q} \quad (16)$$

It is only necessary to elaborate the proof for the case $\|g\|_{L^q} > 0$.

Let us consider the function f defined on the set E by means of the formula

$$f(x) = |g(x)|^{q-1} \operatorname{sgn} g(x) *$$

Obviously, the function f is measurable and $|f|^{\frac{q}{q-1}}$ is summable. But we have $\frac{q}{q-1} = p$ and consequently $f \in L^p$, the norm of f in this space being

$$\|f\|_{L^p} = \left(\int_E |g|^q d\mu \right)^{\frac{1}{p}} = \|g\|_{L^q}^{\frac{q}{p}} \quad (17)$$

Furthermore,

$$F(f) = \int_E |g|^q d\mu = \|g\|_{L^q}^q$$

whence, by equality (17), it follows that

$$F(f) = \|f\|_{L^p} \|g\|_{L^q}^{q - \frac{q}{p}}$$

But $q - \frac{q}{p} = 1$, and hence

$$F(f) = \|f\|_{L^p} \|g\|_{L^q}$$

The last equality implies the relation $\|F\| \geq \|g\|_{L^q}$ which, together with the (above) opposite inequality, results in equality (16).

F. Riesz proved in 1910 that for the case when $E \subset R_1$ formula (15) gives the general representation of a linear functional on L^p . Namely, every linear functional F on the space L_E^p ($E \subset R_1$) is expressible by means of formula (15) with an appropriately chosen function $g \in L_E^q$. The function g is specified by the given functional F uniquely to within the equivalence of functions belonging to L_E^q . In other

* For any real number α the function $\operatorname{sgn} \alpha$ (called the *signum function*) is given by the formula

$$\operatorname{sgn} \alpha = \begin{cases} 1 & \text{for } \alpha > 0 \\ -1 & \text{for } \alpha < 0 \\ 0 & \text{for } \alpha = 0 \end{cases}$$

words, for every F there exists a unique element g of the space L^q_E such that

$$F(f) = \int_E fg \, d\mu$$

This theorem of F. Riesz can also be extended to the general case and Theorem IX.4.1 on the representation of a linear functional in the space L^2 can be obtained as its special case.

The Radon Integral

§ 1. Variations of Additive Set Functions

In this chapter we shall consider a generalization of the Lebesgue integral suggested by J. Radon (1887-1956), an Austrian mathematician. To begin with, we shall discuss some further properties of additive set functions (see Sec. IV.1). Throughout this chapter, when speaking of finitely additive set functions, we shall simply call them *additive*.

Let X be an abstract set and $\mathfrak{S}$ an algebra of some of its subsets. We shall consider an additive set function φ defined for all $A \in \mathfrak{S}$. As in Chapter IV, we shall agree that the infinite values of φ (provided there are such) can only be of one sign. We shall also assume that $\varphi\emptyset = 0$.

Definition. For the given function φ , the set functions expressed in terms of φ by means of the formulas

$$\varphi_+(A) = \sup_{B \subset A, B \in \mathfrak{S}} \varphi(B), \quad \varphi_-(A) = \sup_{B \subset A, B \in \mathfrak{S}} \{ -\varphi(B) \}$$

and

$$|\varphi|(A) = \varphi_+(A) + \varphi_-(A) \quad (A \in \mathfrak{S})$$

are called, respectively, the *positive variation*, the *negative variation* and the *total variation* of φ .*

It should be noted that if the function φ is *bounded*, that is if there exists a constant K such that $|\varphi(A)| \leq K$ for all $A \in \mathfrak{S}$, all the three variations of φ have finite values. If φ is only bounded above (below) the variation φ_+ (φ_-) is finite. The converses of these assertions are also true.**

* As will be shown, the expression $|\varphi|(A)$ makes sense for any $A \in \mathfrak{S}$.

** Any nonnegative additive set function, defined on an algebra, which only assumes finite values is necessarily bounded. This follows from its monotonicity.

Theorem XI.1.1. *All the three variations φ_+ , φ_- and $|\varphi|$ are nonnegative additive set functions on the algebra $\mathfrak{S}$. If $\varphi(A)$ is finite for a set $A \in \mathfrak{S}$ then*

$$|\varphi|(A) = \sup_{\substack{B_1, B_2 \in \mathfrak{S} \\ B_1 \cup B_2 = A}} \{\varphi(B_1) - \varphi(B_2)\} \quad (1)$$

Proof. Since the system of all the subsets $B \subset A$ ($A, B \in \mathfrak{S}$) includes the empty set $B = \emptyset$ and $\varphi(\emptyset) = 0$ we must have $\varphi_+(A) \geq 0$ for any $A \in \mathfrak{S}$. Moreover, it is clear that the function φ_+ is monotone: if $B \subset A$ then $\varphi_+(B) \leq \varphi_+(A)$.

Let A_1 and A_2 be two disjoint sets belonging to $\mathfrak{S}$ and let $A = A_1 \cup A_2$. We shall prove that

$$\varphi_+(A) = \varphi_+(A_1) + \varphi_+(A_2) \quad (2)$$

If at least one of the values $\varphi_+(A_1)$ and $\varphi_+(A_2)$ is equal to $+\infty$ then, by the monotonicity of φ_+ , we also have $\varphi_+(A) = +\infty$; in this case equality (2) is obvious. Therefore we shall suppose that both $\varphi_+(A_1)$ and $\varphi_+(A_2)$ are finite.

Taking an arbitrary $B \subset A$ ($B \in \mathfrak{S}$) and putting $B_1 = B \cap A_1$ and $B_2 = B \cap A_2$ we can write $B = B_1 \cup B_2$ and

$$\varphi(B) = \varphi(B_1) + \varphi(B_2) \leq \varphi_+(A_1) + \varphi_+(A_2)$$

On passing to the supremum in the left member we obtain

$$\varphi_+(A) \leq \varphi_+(A_1) + \varphi_+(A_2) \quad (3)$$

On the other hand, given an arbitrary $\varepsilon > 0$, there are some sets $B_1 \subset A_1$ and $B_2 \subset A_2$ ($B_1, B_2 \in \mathfrak{S}$) such that

$$\varphi(B_1) > \varphi_+(A_1) - \frac{\varepsilon}{2} \text{ and } \varphi(B_2) > \varphi_+(A_2) - \frac{\varepsilon}{2}$$

Then $B = B_1 \cup B_2 \subset A$ and

$$\varphi(B) > \varphi_+(A_1) + \varphi_+(A_2) - \varepsilon$$

Consequently, we also have

$$\varphi_+(A) > \varphi_+(A_1) + \varphi_+(A_2) - \varepsilon$$

By the arbitrariness of ε , it follows that

$$\varphi_+(A) \geq \varphi_+(A_1) + \varphi_+(A_2) \quad (4)$$

Inequalities (3) and (4) imply equality (2), and hence the additivity of the function φ_+ has been proved.

The definition of the function φ_- shows that it is the positive variation of the function $-\varphi$. Since $-\varphi$ is also an additive function we see that, according to what has been proved above, the variation φ_- is a nonnegative additive function. The sum of two additive functions is obviously additive, and therefore the total variation is also an additive set function.

It only remains to prove formula (1). If $\varphi(A)$ is finite $\varphi(B)$ is also finite for any $B \subset A$ ($B \in \mathfrak{S}$) (see Sec. IV.1). Therefore the difference on the right-hand side of formula (1) makes sense. For arbitrary $B_1, B_2 \subset A$ ($B_1, B_2 \in \mathfrak{S}$) we have

$$\varphi(B_1) - \varphi(B_2) \leq \varphi_+(A) + \varphi_-(A) = |\varphi|(A) \quad (5)$$

On the other hand, given arbitrary numbers $l_1 < \varphi_+(A)$ and $l_2 < \varphi_-(A)$, there are $B_1, B_2 \subset A$ ($B_1, B_2 \in \mathfrak{S}$) such that

$$\varphi(B_1) > l_1 \quad \text{and} \quad -\varphi(B_2) > l_2$$

Then

$$\varphi(B_1) - \varphi(B_2) > l_1 + l_2$$

Comparing the last inequality with (5) and taking into account the arbitrariness of l_1 and l_2 we arrive at formula (1).

Theorem XI.1.2. *Every additive set function φ , defined on the algebra $\mathfrak{S}$, bounded above or below is representable as the difference of two nonnegative additive set functions.*

Proof. To prove the theorem we shall derive the equality

$$\varphi = \varphi_+ - \varphi_- \quad (6)$$

Relation (6) is referred to as the *Jordan decomposition* of φ . We shall elaborate the proof for the case when the function φ is bounded below (the other case is treated similarly).

Since under the conditions of the theorem the function φ_- is finite equality (6) is equivalent to the relation $\varphi_+(A) = \varphi(A) + \varphi_-(A)$ for any $A \in \mathfrak{S}$. If $\varphi(A) = +\infty$ then also $\varphi_+(A) = +\infty$, and the last relation is evident. In

the case $\varphi(A) < +\infty$ we have

$$\begin{aligned}\varphi(A) + \varphi_-(A) &= \varphi(A) \sup_{B \subset A, B \in \mathfrak{S}} \{-\varphi(B)\} = \\ &= \sup_{B \subset A, B \in \mathfrak{S}} \{\varphi(A) - \varphi(B)\} = \sup_{B \subset A, B \in \mathfrak{S}} \{\varphi(A \setminus B)\} = \\ &= \sup_{B' \subset A, B' \in \mathfrak{S}} \{\varphi(B')\} = \varphi_+(A)^*\end{aligned}$$

Next, we shall show that *the total variation of the function φ can also be determined with the aid of the formula*

$$|\varphi|(A) = \sup \sum_{i=1}^n |\varphi(A_i)| \quad (7)$$

where the supremum is taken over all finite partitions of the set A into disjoint sets $A_i \in \mathfrak{S}$.

Indeed, if $\varphi(A) = \pm\infty$ then $|\varphi|(A) = +\infty$, and for each partition at least one of the values $\varphi(A_i)$ is equal to plus or minus infinity; in this case equality (7) is trivial.

In what follows we shall assume that $\varphi(A)$ is finite. Then taking an arbitrary partition of the set A into a finite number of disjoint sets A_i we combine in one group the sets A_i for which $\varphi(A_i) > 0$ and in the other group those A_i 's for which $\varphi(A_i) < 0$. Let B_1 denote the union of the sets in the first group and B_2 the union of the sets in the second group. By formula (1)

$$\sum_{i=1}^n |\varphi(A_i)| = \varphi(B_1) - \varphi(B_2) \leq |\varphi|(A)$$

Consequently, if M denotes the right-hand side of formula (7) then also

$$M \leq |\varphi|(A) \quad (8)$$

On the other hand, for an arbitrary number $l < |\varphi|(A)$ there are subsets $B_1, B_2 \subset A$ ($B_1, B_2 \in \mathfrak{S}$) such that

$$\varphi(B_1) - \varphi(B_2) > l \quad (9)$$

Let us put $B = B_1 \cap B_2$, $A_1 = B_1 \setminus B$ and $A_2 = B_2 \setminus B$. All the three sets belong to $\mathfrak{S}$, $A_1 \cap A_2 = \emptyset$ and

$$\varphi(B_1) = \varphi(B) + \varphi(A_1), \quad \varphi(B_2) = \varphi(B) + \varphi(A_2) \quad (10)$$

* When B runs through all the subsets of A belonging to $\mathfrak{S}$ the set $B' = A \setminus B$ also runs through all the subsets of A belonging to $\mathfrak{S}$.

Taking the set $A_3 = A \setminus (A_1 \cup A_2)$ we obtain from (9) and (10) the inequality

$$l < \varphi(A_1) - \varphi(A_2) \leq |\varphi(A_1)| + |\varphi(A_2)| + |\varphi(A_3)| \leq M$$

By the arbitrariness of l , it now follows that the inequality opposite to (8) takes place, and consequently equality (7) holds.

Formula (7) indicates another approach to the definition of the positive, negative and total variations of a *bounded* additive set function. In this case we can first define the total variation by means of formula (7) and prove its additivity with the aid of some simple considerations. Then the positive and negative variations can be introduced according to the formulas

$$\varphi_+ = \frac{1}{2}(|\varphi| + \varphi) \text{ and } \varphi_- = \frac{1}{2}(|\varphi| - \varphi)$$

Theorem XI.1.3. *If a set function φ is σ -additive on the algebra $\mathfrak{S}$ all the three variations φ_+ , φ_- and $|\varphi|$ are also σ -additive.*

Proof. Let $A = \bigcup_{i=1}^{\infty} A_i$ where A and all A_i 's belong to $\mathfrak{S}$ and the sets A_i are disjoint. The finite additivity and the monotonicity of φ_+ imply that

$$\sum_{i=1}^k \varphi_+(A_i) = \varphi_+\left(\bigcup_{i=1}^k A_i\right) \leq \varphi_+(A)$$

for any k , and therefore

$$\sum_{i=1}^{\infty} \varphi_+(A_i) \leq \varphi_+(A) \quad (11)$$

Let us take an arbitrary set $B \subset A$ ($B \in \mathfrak{S}$) and put $B_i = B \cap A_i$ ($i = 1, 2, \dots$). Then $B = \bigcup_{i=1}^{\infty} B_i$, and the sets B_i belong to $\mathfrak{S}$ and are disjoint. It follows that

$$\varphi(B) = \sum_{i=1}^{\infty} \varphi(B_i) \leq \sum_{i=1}^{\infty} \varphi_+(A_i)$$

Taking the supremum of the left-hand member we see that

$$\varphi_+(A) \leq \sum_{i=1}^{\infty} \varphi_+(A_i) \quad (12)$$

The σ -additivity of the function φ_+ now follows from (11) and (12). Incidentally we have proved that the function φ_- is also σ -additive since it coincides with the positive variation of $-\varphi$. Therefore the function $|\varphi| = \varphi_+ + \varphi_-$ is also σ -additive.

In the two theorems below we assume that the domain of definition $\mathfrak{S}$ of φ is a σ -algebra.

Theorem XI.1.4. *If the function φ takes finite values and is σ -additive on the σ -algebra $\mathfrak{S}$ it is bounded.*

Proof. Let us suppose that

$$\sup_{A \subset X, A \in \mathfrak{S}} |\varphi(A)| = +\infty$$

and put $A_1 = X$. Then there exists a subset $B \subset A_1$ such that $|\varphi(B)| > |\varphi(A_1)| + 1$. This, in its turn, implies that

$$|\varphi(A_1 \setminus B)| \geq |\varphi(B)| - |\varphi(A_1)| > 1$$

One of the sets B and $A_1 \setminus B$ is such that the function φ is unbounded on the collection of its subsets (belonging to $\mathfrak{S}$). Let A_2 denote that set. Then

$$\sup_{A \subset A_2, A \in \mathfrak{S}} |\varphi(A)| = +\infty$$

and $|\varphi(A_1) - \varphi(A_2)| > 1$.

Proceeding in this way we construct a decreasing infinite sequence of sets $A_n \in \mathfrak{S}$ such that

$$|\varphi(A_n) - \varphi(A_{n+1})| > 1$$

for every $n = 1, 2, \dots$. Since $\mathfrak{S}$ is a σ -algebra we have

$A = \bigcap_{n=1}^{\infty} A_n \in \mathfrak{S}$ (see Sec. 1.7), and by Theorem IV.1.2,

$\varphi(A_n) \rightarrow \varphi(A)$. But this contradicts the foregoing inequality, which proves the theorem.

Corollary. *If a σ -additive set function φ defined on a σ -algebra $\mathfrak{S}$ is σ -finite* its total variation $|\varphi|$ and, together with it, the variations φ_+ and φ_- are also σ -finite.*

Indeed, if φ is σ -finite the above theorem indicates that the space X is representable as a countable union of sets X_n such that, for every n , the function φ is bounded on the part of the σ -algebra $\mathfrak{S}$ which is contained in X_n .

Theorem XI.1.5. *If a σ -additive function φ defined on a σ -algebra $\mathfrak{S}$ does not assume the value $+\infty$ ($-\infty$) it is bounded above (below).*

Proof. We shall prove the theorem by contradiction. If we suppose that a σ -additive function φ does not take the value $+\infty$ but is not bounded above then there must exist a set $A_1 \in \mathfrak{S}$ such that $\varphi(A_1) > 1$. Since $\varphi(A_1) \neq +\infty$ the value $\varphi(A_1)$ is finite and, consequently, $\varphi(A)$ is also finite for any $A \subset A_1$ ($A \in \mathfrak{S}$). Therefore, by the foregoing theorem, φ is bounded on the part of the σ -algebra $\mathfrak{S}$ contained in A_1 . This means that φ cannot be bounded above on the part of the σ -algebra $\mathfrak{S}$ which is contained in $A \setminus A_1$. Consequently, there exists a set $A_2 \subset A \setminus A_1$ ($A_2 \in \mathfrak{S}$) such that $\varphi(A_2) > 1$. Now we can analogously find $A_3 \subset A \setminus (A_1 \cup A_2)$ ($A_3 \in \mathfrak{S}$) for which $\varphi(A_3) > 1$. Proceeding in this way indefinitely we obtain a sequence of disjoint sets $A_i \in \mathfrak{S}$ ($i = 1, 2, \dots$) on each of which $\varphi(A_i) > 1$. Let us put $B = \bigcup_{i=1}^{\infty} A_i$. Since $\mathfrak{S}$ is a σ -algebra, B

must belong to $\mathfrak{S}$, and $\varphi(B) = \sum_{i=1}^{\infty} \varphi(A_i) = +\infty$, which contradicts the hypothesis. The assumption that φ is not bounded above has thus been disproved. The case of a function not assuming the value $-\infty$ is treated similarly.

Let us consider some examples. (i) Let μ be a σ -finite measure in X defined on a σ -algebra $\mathfrak{S}$. Suppose that there is a measurable function g , defined and finite almost everywhere on a measurable set $E \subset X$, such that the integral

* For an arbitrary set function φ defined on an algebra $\mathfrak{S}$ the notion of σ -finiteness has the same meaning as in the case of a measure:

there exists a representation $X = \bigcup_{n=1}^{\infty} X_n$ where $X_n \in \mathfrak{S}$ such that the values $\varphi(X_n)$ are finite for all $n = 1, 2, \dots$

$\int_E g d\mu$ with respect to the measure μ exists. Let us put

$$\varphi(A) = \int_A g d\mu$$

where A is measurable subset of E . Then the set function φ is σ -additive. Let us show that

$$\left. \begin{aligned} \varphi_+(A) &= \int_A g_+ d\mu, & \varphi_-(A) &= \int_A g_- d\mu \\ |\varphi|(A) &= \int_A |g| d\mu \end{aligned} \right\} \quad (13)$$

(see Sec. VIII.2 and, in particular, formula (4)). It suffices to derive the first of these three formulas since the other two will follow from it.

By the definition of the function g_+ , there is a measurable set $B \subset A$ such that

$$g_+(x) = \begin{cases} g(x) & \text{on } B \\ 0 & \text{on } A \setminus B \end{cases}$$

Then

$$\int_A g_+ d\mu = \int_B g d\mu = \varphi(B) \leq \varphi_+(A)$$

On the other hand, for any measurable set $B \subset A$ we can write

$$\varphi(B) = \int_B g d\mu = \int_B g_+ d\mu - \int_B g_- d\mu \leq \int_B g_+ d\mu \leq \int_A g_+ d\mu$$

and consequently $\varphi_+(A) \leq \int_A g_+ d\mu$. The desired equality now follows from the two opposite inequalities we have established.

It is readily seen that the function φ in this example is σ -finite.

(ii) Now we shall consider a generalization of one of the examples in Sec. VII.1. Let X be an arbitrary infinite set from which a countable subset of points $\{x_i\}$ is chosen

and let $\varphi_i \neq 0$ be real numbers (*charges*) such that the series $\sum_{i=1}^{\infty} \varphi_i$ is absolutely convergent. Taking an arbitrary $E \subset X$ we put

$$\varphi(E) = \sum_{i(x_i \in E)} \varphi_i$$

Thus, $\varphi(E)$ is equal to the sum of the charges φ_i at the points x_i contained in E ; the charges can now be of any sign. It is obvious that φ is a bounded σ -additive set function for which

$$\varphi_+(E) = \sum_{i(x_i \in E)} \varphi_i^+, \quad \varphi_-(E) = \sum_{i(x_i \in E)} \varphi_i^-$$

and

$$|\varphi|(E) = \sum_{i(x_i \in E)} |\varphi_i|$$

where

$$\varphi_i^+ = \begin{cases} \varphi_i & \text{for } \varphi_i > 0 \\ 0 & \text{for } \varphi_i < 0 \end{cases} \quad \text{and} \quad \varphi_i^- = \begin{cases} 0 & \text{for } \varphi_i > 0 \\ |\varphi_i| & \text{for } \varphi_i < 0 \end{cases}$$

(iii) Let F be an arbitrary wide-sense increasing (i.e. nondecreasing) function defined on an interval $I = [a, b] \subset \mathbf{R}_1$. We shall consider the algebra $\mathfrak{S}$ generated by the system of all the cells $\Delta \subset [a, b)$.* This means that each $A \in \mathfrak{S}$ is representable in the form of a finite union of disjoint cells: $A = \bigcup_{i=1}^p \Delta_i$. For each nonempty cell $\Delta = [\alpha, \beta) \subset I$ we put

$$\varphi(\Delta) = F(\beta) - F(\alpha) \quad (14)$$

and for the empty cell we put $\varphi(\emptyset) = 0$. Now, taking an arbitrary set $A \in \mathfrak{S}$ and using the above representation of A as a union of pairwise disjoint cells we define on $\mathfrak{S}$ the set function

$$\varphi(A) = \sum_{i=1}^p \varphi(\Delta_i) \quad (15)$$

It can easily be checked that in this definition $\varphi(A)$ does not in fact depend on the way in which A is represented in

* $\mathfrak{S}$ is an algebra of subsets of the cell $[a, b)$.

the form of a union of disjoint cells and that the function φ is additive. At the same time if the function F is not continuous from the left the function φ is not σ -additive.*

Indeed, suppose that the function F is discontinuous from the left at a point c . Let us take an increasing number sequence $c_n \rightarrow c - 0$ (we assume that $c_1 = a$) and put $\Delta_n = [c_n, c_{n+1})$. Then the cells Δ_n are pairwise disjoint $\bigcup_{n=1}^{\infty} \Delta_n = \Delta$ where $\Delta = [a, c)$,

$$\sum_{n=1}^{\infty} \varphi(\Delta_n) = \sum_{n=1}^{\infty} [F(c_{n+1}) - F(c_n)] = F(c - 0) - F(a)$$

and $\varphi(\Delta) = F(c) - F(a)$. But $F(c) > F(c - 0)$ and therefore the function φ is not σ -additive.

A similar construction can be made for an arbitrary function F of bounded variation defined on the interval $[a, b]$.** The function φ determined, for such F , by formulas (14) and (15) is also a bounded additive set function on $\mathfrak{S}$ but in this case it can assume negative values. From formula (7) it follows immediately that the function $|\varphi|$ coincides with the total variation of the function F , that is, more precisely, for any cell $\Delta = [\alpha, \beta) \subset I$ we have

$$|\varphi|(\Delta) = V_{\alpha}^{\beta}(F)$$

* Conversely, if F is continuous from the left it can be proved that φ is σ -additive. Cf. Sec. XI.6; in particular, see Theorem XI.6.1.

** We remind the reader that a function F defined on an interval $[a, b]$ is said to be of *bounded variation* (on that interval) if there is a nonnegative constant K such that for any partition of the interval $[a, b]$ into a finite number of subintervals by means of points $a = x_0 < x_1 < \dots < x_m = b$ the inequality

$$\sum_{i=0}^{m-1} |F(x_{i+1}) - F(x_i)| \leq K$$

holds. The least of the constants K is called the *total variation* of the function F on $[a, b]$ (denoted $V_a^b(F)$). An important property is that every function of bounded variation is representable as the difference of two increasing functions. The converse is also true which is almost evident. (On the properties of functions of bounded variation see, for instance, A. N. Kolmogorov, S. V. Fomin, *Introductory Real Analysis*, Prentice-Hall, New Jersey, 1970 (the English translation of the Russian original).)

§ 2. The Radon Integral

The basic idea of the generalization of the notion of the integral suggested by J. Radon is that the integration with respect to a measure is replaced by the integration with respect to an arbitrary σ -additive set function (which can assume negative values as well). In its properties the Radon integral turns out to be rather close to the Lebesgue integral.

We shall begin with proving an auxiliary lemma.

Lemma XI.2.1. *Let X be an abstract set, $\mathfrak{S}$ be a σ -algebra of its (measurable) subsets and let μ and ν be two arbitrary σ -finite measures defined on $\mathfrak{S}$. If f is a function defined on a set $E \in \mathfrak{S}$ summable on E with respect to each of the measures it is also summable on E with respect to the sum $\omega = \mu + \nu$ of these measures and*

$$\int_E f d\omega = \int_E f d\mu + \int_E f d\nu \quad (16)$$

Proof. Let us first assume that the function f is measurable and bounded on E and that the measures μE and νE are finite. The Lebesgue-Darboux sums constructed for an arbitrary partition τ of the set E into disjoint measurable subsets e_i ($E = \bigcup_{i=1}^p e_i$) are obviously connected by the relation

$$\begin{aligned} s_\omega(\tau; f) &= \sum_{i=1}^p m_i \omega e_i = \sum_{i=1}^p m_i \mu e_i + \sum_{i=1}^p m_i \nu e_i \leq \\ &\leq \int_E f d\mu + \int_E f d\nu \leq \sum_{i=1}^p M_i \mu e_i + \sum_{i=1}^p M_i \nu e_i = \\ &= \sum_{i=1}^p M_i \omega e_i = S_\omega(\tau; f) \end{aligned}$$

The function f being integrable with respect to the measure ω , the sums $s_\omega(\tau; f)$ and $S_\omega(\tau; f)$ can be made arbitrarily close to each other and, by the above relation, the common value of the supremum of the former and the infimum of the

latter coincides with the sum of the integrals $\int_E f d\mu + \int_E f dv$. This proves equality (16).

Next we suppose that the function f is measurable, bounded and nonnegative on E and that $\omega E = +\infty$. Let us represent E in the form $E = \bigcup_{k=1}^{\infty} E_k$ where the sets E_k are pairwise disjoint and $\omega E_k < +\infty$ for all k .* As has been proved,

$$\int_{E_k} f d\omega = \int_{E_k} f d\mu + \int_{E_k} f dv \quad (k = 1, 2, \dots)$$

On summing these equalities with respect to k we arrive at (16).

Now we shall discard the condition that f is bounded and only consider it measurable and nonnegative. Let us take the truncated functions f_m corresponding to f . By what has been proved, we have

$$\int_E f_m d\omega = \int_E f_m d\mu + \int_E f_m dv$$

Passing to the limit in this equality as $m \rightarrow \infty$ (see Sec. VIII.4) we obtain formula (16).

Finally, in the general case when f can assume negative values as well we represent it as $f = f_+ - f_-$, write down equality (16) for each of the functions f_+ and f_- and then perform the termwise subtraction.

Remark. The above proof shows that for a nonnegative measurable function f equality (16) is valid even without the requirement that it should be summable. Moreover, if μ and ω are two σ -finite measures defined on $\mathfrak{G}$, $\mu A \leq \omega A$, for any $A \subset E$ ($A \in \mathfrak{G}$) and f is nonnegative and measurable then

$$\int_E f d\mu \leq \int_E f d\omega$$

* The σ -finiteness of the measures μ and ν implies that the measure ω is also σ -finite.

Now we shall state the definition of the Radon integral. As before, we denote by X an abstract set and consider a σ -algebra $\mathfrak{S}$ of its subsets regarded, by definition, as measurable and a σ -additive and σ -finite set function φ defined on $\mathfrak{S}$. From the foregoing section it follows (see Theorem XI.1.3 and the corollary of Theorem XI.1.4) that φ_+ and φ_- are σ -finite measures in X . Furthermore, since φ can only assume infinite values of one sign Theorem XI.1.5 indicates that it is bounded from at least one side, and therefore equality (6) applies to it.

Definition. Let f be a measurable function defined on a measurable set $E \subset X$. The Radon integral of the function f with respect to the function φ over the set E is defined by means of the formula

$$\int_E f d\varphi = \int_E f d\varphi_+ - \int_E f d\varphi_- \quad (17)$$

under the assumption that the integrals on the right-hand side of the formula and their difference make sense*. If otherwise, the integral $\int_E f d\varphi$ does not make sense. If

the integral $\int_E f d\varphi$ has a finite value the function f is said to be *Radon summable on E with respect to the function φ* .

In contrast to the integral with respect to a measure, the Radon integral $\int_E f d\varphi$ of a nonnegative measurable

function f may not exist. In particular, the fact that $\int_E f d\varphi$

exists no longer implies that $\int_E |f| d\varphi$ also exists. At the

same time, from the lemma (see also the remark after the lemma) it follows that f is summable with respect to φ if and only if the function f and, together with it, the modulus $|f|$ are summable with respect to the total variation $|\varphi|$.

* The condition that the difference should make sense means that either at least one of the integrals on the right-hand side is finite or one of them is equal to $+\infty$ while the other is equal to $-\infty$.

The Radon integral is sometimes also referred to as the *Lebesgue-Stieltjes** integral.

From the definition of the Radon integral it readily follows that most of the properties of the Lebesgue integral proved in Sec. VIII.3 extend to the Radon integral literally or with slight changes. In particular, the integral remains σ -additive and Theorem VIII.3.8 continues to hold. Formula (17) directly implies the following estimation for the Radon integral:

$$\left| \int_E f d\varphi \right| \leq \int_E |f| d|\varphi| \quad (18)$$

It follows that the Radon integral of a summable function possesses the property of absolute continuity (see Theorem VIII.3.3) with respect to the measure $|\varphi|$. The theorems involving the notion of equivalent functions and relations which hold almost everywhere also continue to hold if the term "almost everywhere" is understood relative to the measure $|\varphi|$: "almost everywhere on E " should mean "for all $x \in E$ with a possible exception of the points of a subset E' for which $|\varphi|(E') = 0$ ".

The following property of the Radon integral complements Lemma XI.2.1. *Let φ and ψ be two σ -additive set functions with finite values defined on $\mathfrak{S}$. If f is a summable function on E both with respect to φ and with respect to ψ it is also summable with respect to $\varphi \pm \psi$ and*

$$\int_E f d(\varphi \pm \psi) = \int_E f d\varphi \pm \int_E f d\psi$$

This assertion can easily be proved on the basis of Lemma XI.2.1, and we leave it to the reader.

For the σ -additive function φ constructed in Example 2 of Sec. XI.1 from given point charges the Radon integral reduces to a sum:

$$\int_E f d\varphi = \sum_{i(x_i \in E)} f(x_i) \varphi_i$$

(also cf. Example 1 in Sec. VII.1).

* T. J. Steiltjes (1856-1894), a Dutch mathematician, who suggested a notion of the integral which is a direct generalization of the Riemann integral (see Sec. XI.5).

As another example, let us consider the Radon integral with respect to a set function which itself is representable as the Lebesgue integral of a point function. Let μ be a σ -finite measure in a space X defined on a σ -algebra $\mathfrak{S}$ and let f and g be two measurable functions defined and finite almost everywhere on a set $E \in \mathfrak{S}$. Let the integral $\int_E g d\mu$ exist and φ be the σ -additive and σ -finite function defined for any measurable set $A \subset E$ by the formula

$$\varphi(A) = \int_A g d\mu$$

(Ex. 1, Sec. XI.1). We shall prove the following theorem.

Theorem XI.2.1. *For the function f to be summable on the set E with respect to the set function φ it is necessary and sufficient that the product fg be summable on E with respect to the measure μ and*

$$\int_E f d\varphi = \int_E fg d\mu \quad (19)$$

This theorem shows that the Radon integral of the function f with respect to the function φ reduces, in the case under consideration, to the Lebesgue integral of the function fg with respect to the measure μ .

Proof. We shall begin with the case when f and g are nonnegative (then φ is also nonnegative, that is φ is a measure). Let us also suppose that $\varphi E < +\infty$.

If f is bounded we take an arbitrary partition τ of the set E into pairwise disjoint measurable subsets e_i ($i = 1, 2, \dots, p$) and consider the corresponding Lebesgue-Darboux sums. For these sums we have the obvious relation

$$\begin{aligned} s_\varphi(\tau; f) &= \sum_{i=1}^p m_i \varphi e_i = \sum_{i=1}^p m_i \int_{e_i} g d\mu \leq \\ &\leq \sum_{i=1}^p \int_{e_i} fg d\mu = \int_E fg d\mu \leq \sum_{i=1}^p M_i \int_{e_i} g d\mu = \\ &= \sum_{i=1}^p M_i \varphi e_i = S_\varphi(\tau; f) \end{aligned}$$

The integral $\int_E f d\varphi$ exists and its value lies between $s_\varphi(\tau; f)$ and $S_\varphi(\tau; f)$, and hence these sums can be made arbitrarily close to each other whence directly follows equality (19).

The case of an unbounded function f can be treated in an obvious way by using the truncated functions and Levi's theorem.

Now suppose that $\varphi E = +\infty$. In this case we partition E into disjoint sets E_i ($i = 1, 2, \dots$) with $\varphi E_i < +\infty$, write down equality (19) for the integrals over the sets E_i and then add these equalities together.

In the general case when f and g may assume values of different signs we use the functions f_+, f_-, g_+, g_- , φ_+ and φ_- . By what has been proved for nonnegative functions taking into account the formulas for φ_+ and φ_- from Example 1 in Sec. XI.1 we obtain

$$\int_E f_+ d\varphi_+ = \int_E f_+ g_+ d\mu, \quad \int_E f_+ d\varphi_- = \int_E f_+ g_- d\mu,$$

$$\int_E f_- d\varphi_+ = \int_E f_- g_+ d\mu$$

and

$$\int_E f_- d\varphi_- = \int_E f_- g_- d\mu$$

whence

$$\int_E f d\varphi_+ = \int_E f g_+ d\mu \quad \text{and} \quad \int_E f d\varphi_- = \int_E f g_- d\mu$$

This leads to equality (19) provided that one of the integrals in (19) has a finite value.*

Remark. The same argument can be applied to prove that for the integral $\int_E f d\varphi$ to exist under the conditions of Theorem XI.2.1 it is necessary and sufficient that the integral $\int_E f g d\mu$ exist; then there also holds equality (19).

* In this case all the integrals written above also have finite values.

§ 3. Integration with Respect to a Finitely Additive Set Function

The next step in the generalization of the notion of the integral is to pass to the integration with respect to a finitely additive set function. The Soviet mathematicians G. M. Fikhtengolts and L. V. Kantorovich and the American mathematician T. H. Hildebrandt were the first to study such an integral. Here we shall limit ourselves to the case when both the integrand function and the function with respect to which the integration is carried out are bounded.

Thus, let X be an abstract set and $\mathfrak{S}$ be an algebra of some of its subsets. We shall consider a bounded additive set function φ defined on $\mathfrak{S}$. Let us begin with the case when φ is nonnegative.

Taking a real point function f defined and bounded on a set $E \in \mathfrak{S}$ and an arbitrary partition τ of the set E into a finite number of disjoint sets $e_i \in \mathfrak{S}$ ($i = 1, 2, \dots, p$) we form the Lebesgue-Darboux sums

$$s(\tau; f) = \sum_{i=1}^p m_i \varphi(e_i) \text{ and } S(\tau; f) = \sum_{i=1}^p M_i \varphi(e_i)$$

where

$$m_i = \inf_{x \in e_i} f(x) \text{ and } M_i = \sup_{x \in e_i} f(x) \quad (i = 1, 2, \dots, p)$$

The properties of the Lebesgue-Darboux sums established in Sec. VII.1 continue to hold in the case when the function φ is finitely additive since the σ -additivity of the measure was not used in the course of the proof of these properties.

We shall define the integral $\int_E f d\varphi$ by analogy with the Lebesgue integral (Sec. VII.1), that is as the common value of $\sup_{\tau} s(\tau; f)$ and $\inf_{\tau} S(\tau; f)$ in case the supremum and the infimum coincide.

In the general case when φ can assume values of any sign the integral $\int_E f d\varphi$ is defined by means of formula (17) under the assumption that both integrals on its right-hand side exist.

As in Sec. VII.1, it can be proved that if a bounded function f is $\mathfrak{S}$ -measurable* it is integrable. But in this case the class of integrable functions can be essentially wider than that of measurable functions.

The integral with respect to a finitely additive set function has much in common with the Radon integral. It can easily be shown that if f is φ -integrable** then f and $|f|$ are $|\varphi|$ -integrable ($|\varphi|$ is the total variation of φ) and inequality (18) holds. Furthermore, this inequality implies that if $|f(x)| \leq K$ then

$$\left| \int_E f d\varphi \right| \leq K |\varphi|(E) \quad (20)$$

Indeed, by definition,

$$\int_E f d\varphi = \int_E f d\varphi_+ - \int_E f d\varphi_-$$

A theorem analogous to Theorem VII.2.1 shows that

$$\left| \int_E f d\varphi_+ \right| \leq K \varphi_+(E) \quad \text{and} \quad \left| \int_E f d\varphi_- \right| \leq K \varphi_-(E)$$

These formulas directly imply inequality (20).

Instead of the σ -additivity, in this case we can only prove the *finite* additivity of the integral. For bounded φ -integrable functions f and g there also holds Theorem VII.2.3. In this connection it should be noted that when proving formula (5), Chapter VII, we can no longer deduce the existence of the integral $\int_E (f + g) d\varphi$ from the measurability of the function $f + g$ because the $\mathfrak{S}$ -measurability of f and g is not supposed here***. But the course of the proof of formula (5) does in fact show that $f + g$ is integrable.

Here we shall not dwell on further properties of the integral.

Let us consider the special case when $X = \mathbf{R}_n$, the algebra $\mathfrak{S}$ contains all closed sets, E is a bounded closed

* As usual, the $\mathfrak{S}$ -measurability of f means that all the Lebesgue sets of the function f are contained in the algebra $\mathfrak{S}$ (see Sec. VI.1).

** I.e. integrable with respect to φ .

*** Even if we supposed f and g to be $\mathfrak{S}$ -measurable this would not imply the measurability of $f + g$.

set belonging to R_n and f is continuous on E . We shall show directly that in this case the integral $\int_E f d\varphi$ exists for any bounded additive set function φ on $\mathfrak{S}$.

It is sufficient to carry out the proof for the case of a nonnegative function φ . Given an arbitrary $\varepsilon > 0$, there is $\delta > 0$ such that the oscillation of the function f on any subset of E with diameter less than δ is less than ε (since the function f is uniformly continuous). Let us partition the set E into a finite number of disjoint sets e_i by means of a network of cells with diameters less than δ^* (each set e_i is the intersection of the set E with one of the cells). Each cell as a difference of two closed sets belongs to the algebra $\mathfrak{S}$, and therefore $e_i \in \mathfrak{S}$. It now becomes clear that for the partition thus constructed the difference between the upper and lower Lebesgue-Darboux sums is less than $\varepsilon\varphi(E)$. We see that this difference can be made arbitrarily small, and consequently the function f is integrable.

The same result can be obtained in the case when $\mathfrak{S}$ is the algebra of subsets of R_n generated by the system of all cells. If $E \subset R_n$ is a cell with finite edges and the function f is continuous on the closure $\bar{E}$ of E then it is integrable on E with respect to any bounded additive set function φ defined on $\mathfrak{S}$. At the same time, the class of $\mathfrak{S}$ -measurable functions on E only consists of the "piecewise constant" functions, with at most countable number of different values, whose all Lebesgue sets are representable as finite unions of cells.

§ 4. Linear Functionals on the Space of Bounded Functions

The notion of the integral with respect to a finitely additive set function can be applied to derive an analytical representation for linear functionals defined on spaces of bounded functions. The Soviet mathematicians G. M. Fikhtengolts and L. V. Kantorovich were the first to obtain the general representation of a linear functional on a space

* For the diameter of an n -dimensional cell to be less than δ it is sufficient that its every edge be less than $\frac{\delta}{\sqrt{n}}$.

of bounded measurable functions. Theorems below are essentially the result of their studies.

As in Ses. XI.3, let X be an abstract set. In this section by $\mathfrak{S}$ we shall mean a σ -algebra of some of the subsets of X called measurable. The collection of all bounded measurable functions f defined on X will be denoted by B (or, more precisely, $B(X, \mathfrak{S})$). Since $\mathfrak{S}$ is a σ -algebra it follows from the results of Chapter VI (see Sec. VI.2) that B is a linear space. Let us define a norm in B by putting

$$\|f\| = \sup_{x \in X} |f(x)|$$

It is obvious that $\|f\|$ satisfies all the axioms of norm and thus B equipped with this norm is a normed space. The convergence in norm in the space B coincides with the uniform convergence. Finally, since the class of measurable functions is closed with respect to the operation of passing to the limit (for the pointwise convergence see Theorem VI.3.1), B is a complete space.

Theorem XI.4.1. *The general representation of a linear functional on the space B is given by the integral*

$$F(f) = \int_X f d\varphi \quad (21)$$

where φ is an arbitrary bounded additive set function defined on $\mathfrak{S}$, and the norm of the functional is determined by the formula

$$\|F\| = |\varphi|(X) \quad (22)$$

Proof. As was indicated in Sec. XI.3, integral (21) exists for any $f \in B$ and for any function φ satisfying the given conditions. By the properties of this integral, the functional F determined by formula (21) is distributive. From inequality (20) it follows that

$$|F(f)| \leq \|f\| \cdot |\varphi|(X)$$

for any $f \in B$. Consequently, F is a linear functional and

$$\|F\| \leq |\varphi|(X) \quad (23)$$

Conversely, let F be an arbitrary linear functional on the space B . For an arbitrary $A \in \mathfrak{S}$ we now put

$$\varphi(A) = F(f_A)$$

where f_A is the characteristic function of the set A . The additivity of the functional F immediately implies the additivity of the function φ . Furthermore,

$$|\varphi(A)| \leq \|F\| \|f_A\| \leq \|F\|$$

and therefore the function φ is bounded. Hence, the integral in formula (21) exists for any $f \in B$.

Let $f \in B$ and $C < f(x) < D$. On partitioning the interval (C, D) by means of points $C = \lambda_0 < \lambda_1 < \dots < \lambda_m = D$ and forming the sets

$$e_i = X[\lambda_{i-1} \leq f(x) < \lambda_i] \quad (i = 1, 2, \dots, m)$$

we construct the function

$$g = \sum_{i=1}^m \lambda_i f_i$$

where f_i ($i = 1, 2, \dots, m$) is the characteristic function of the set e_i . It is obvious that the function g is measurable and hence $g \in B$. If $x \in e_i$ then

$$|f(x) - g(x)| = |f(x) - \lambda_i| \leq \delta$$

where $\delta = \max(\lambda_i - \lambda_{i-1})$, whence

$$\sup_{x \in X} |f(x) - g(x)| \leq \delta$$

that is $\|f - g\| \leq \delta$. Consequently,

$$|F(f) - F(g)| \leq \|F\| \delta \quad (24)$$

On the other hand,

$$F(g) = \sum_{i=1}^m \lambda_i F(f_i) = \sum_{i=1}^m \lambda_i \varphi(e_i) \quad (25)$$

and

$$\int_X f d\varphi - \sum_{i=1}^m \lambda_i \varphi(e_i) = \int_X (f - g) d\varphi \leq \delta |\varphi|(X) \quad (26)$$

Comparing (24), (25) and (26) we see that δ can be chosen so that the difference

$$F(f) - \int_X f d\varphi$$

becomes arbitrarily small, which proves formula (21).

It now remains to derive equality (22). Given $\varepsilon > 0$, formula (1) shows that there exist some sets $B_1, B_2 \in \mathfrak{S}$ such that

$$\varphi(B_1) - \varphi(B_2) > |\varphi|(X) - \varepsilon$$

Let f_1 and f_2 be, respectively, the characteristic functions of these sets, and $f = f_1 - f_2$. Then f can only assume the values 1, -1 and 0 and therefore $\|f\| \leq 1$. But we have

$$F(f) = F(f_1) - F(f_2) = \varphi(B_1) - \varphi(B_2)$$

and thus

$$\|F\| \geq F(f) > |\varphi|(X) - \varepsilon$$

By the arbitrariness of ε , it follows that $\|F\| \geq |\varphi|(X)$, which, together with (23), implies equality (22).

The theorem has been completely proved. Note that the definition of the space $\mathbf{B}$ does not involve the requirement that there should be defined a measure on the σ -algebra $\mathfrak{S}$.

Remark. It can easily be checked that a given linear functional uniquely specifies the bounded additive set function φ entering in (21). Indeed, if a linear functional F is representable by formula (21) we necessarily have $F(f_A) = \varphi(A)$ for the characteristic function f_A of any set $A \in \mathfrak{S}$, that is the function φ coincides with the one constructed in the course of the proof of the theorem.

We shall also consider a slightly different space of bounded measurable functions on X . Suppose now that there is a measure μ defined on the σ -algebra $\mathfrak{S}$. Let us identify mutually equivalent functions and define the norm by the formula

$$\|f\| = \text{vrai sup } |f(x)|$$

(see the notation in Sec. VIII.5). We thus arrive at a new Banach space which will be denoted as $\mathbf{M}$.

Theorem XI.4.2. *The general representation of a linear functional on the space $\mathbf{M}$ is given by formula (21) in which φ is an arbitrary bounded additive set function, defined for all measurable subsets of X , satisfying the additional condition that $\varphi(A) = 0$ if $\mu A = 0$. The norm of F is expressed by formula (22).*

Proof. Let F be a linear functional on $\mathbf{M}$. As above, we define the function φ on the system of measurable sets by

putting

$$\varphi(A) = F(f_A)$$

and readily see that φ is a bounded additive set function and $\varphi(A) = 0$ for all A with $\mu A = 0$. Now, repeating the foregoing argument, we conclude that the functional F is representable by formula (21) and that $\|F\| \geq |\varphi|(X)$.

Conversely, let φ be an arbitrary set function satisfying the conditions of the theorem. Note that if $\mu A = 0$ then $|\varphi|(A) = 0$. Integral (21) is sure to exist for all $f \in M$. If $f \sim 0$ the space X can be represented as the union of two nonintersecting measurable sets E_1 and E_2 such that $f(x) \equiv 0$ on E_1 and $\mu E_2 = 0$. Then

$$\int_{E_1} f d\varphi = 0$$

On the other hand,

$$\left| \int_{E_2} f d\varphi \right| \leq K |\varphi|(E_2) = 0$$

where K is an upper bound of $|f(x)|$. Consequently,

$$\int_X f d\varphi = 0$$

If $f, g \in M$ and $f \sim g$ we have

$$\int_X f d\varphi - \int_X g d\varphi = \int_X (f - g) d\varphi = 0$$

that is the integrals of type (21) with mutually equivalent integrand functions coincide. Therefore formula (21) specifies a distributive functional on M .

For any $f \in M$, let us put

$$A_1 = X [|f(x)| \leq \|f\|] \quad \text{and} \quad A_2 = X \setminus A_1$$

then $\mu A_2 = 0$ and, consequently, $\int_{A_2} f d\varphi = 0$ while

$$\int_X f d\varphi = \int_{A_1} f d\varphi$$

It follows immediately that

$$\left| \int_{\mathbf{X}} f d\varphi \right| \leq \|f\| \cdot |\varphi|(\mathbf{X}) = \|f\| \cdot |\varphi|(\mathbf{X})$$

Thus, the functional F determined by formula (21) is linear and $\|F\| \leq |\varphi|(\mathbf{X})$. The two opposite inequalities we have obtained result in equality (22), which completes the proof of the theorem.

We can also interpret M as the space of all functions which are measurable and bounded *almost everywhere* on $\mathbf{X}$. For Theorem XI.4.2 to continue to hold for this interpretation it is sufficient to agree that the integral of a function belonging to this class is understood as that of an equivalent measurable function bounded throughout the set $\mathbf{X}$.

§ 5. The Stieltjes Integral on the Line

The *Stieltjes integral* (also called the *Riemann-Stieltjes integral*) is a direct generalization of the classical Riemann integral. Here we shall state its definition in a form which is similar to the definitions of other types of integral studied in this course.

Let us consider two functions defined on an interval $[a, b]$: an arbitrary bounded function f and a wide-sense increasing (i.e. nondecreasing) function g . Taking an arbitrary partition τ of the interval $[a, b]$ into a finite number of subintervals with the aid of points of division $a = x_0 < x_1 < \dots < x_p = b$ we put

$$M_i = \sup_{x \in [x_{i-1}, x_i]} f(x), \quad m_i = \inf_{x \in [x_{i-1}, x_i]} f(x) \\ (i = 1, 2, \dots, p)$$

and form the *Stieltjes-Darboux sums*

$$S(\tau) = \sum_{i=1}^p M_i [g(x_i) - g(x_{i-1})]$$

and

$$s(\tau) = \sum_{i=1}^p m_i [g(x_i) - g(x_{i-1})]$$

As in the case of classical Darboux's sums, it can easily be shown that

$$\sup_{\tau} s(\tau) \leq \inf_{\tau} S(\tau) \quad (27)$$

If the supremum of $s(\tau)$ and the infimum of $S(\tau)$ coincide we call their common value *the Stieltjes integral of the function f with respect to the generating function g* and denote it as

$$\int_a^b f dg \quad \text{or} \quad \int_a^b f(x) dg(x)$$

It should be noted that if g and h are two increasing functions on the interval $[a, b]$ and if both integrals

$\int_a^b f dg$ and $\int_a^b f dh$ exist the integral $\int_a^b f d(g + h)$ also exists and

$$\int_a^b f d(g + h) = \int_a^b f dg + \int_a^b f dh \quad (28)$$

The proof of this assertion is quite simple.

Now suppose that g is an arbitrary function of bounded variation on the interval $[a, b]$.* It can be represented as a difference of two increasing functions: $g = g_1 - g_2$. If among these representations there is at least one for which

both integrals $\int_a^b f dg_1$ and $\int_a^b f dg_2$ exist we put, by definition,

$$\int_a^b f dg = \int_a^b f dg_1 - \int_a^b f dg_2$$

Let us check that the integral $\int_a^b f dg$ thus defined is in fact independent of the particular choice of g_1 and g_2 . If $g = h_1 - h_2$ is another representation of the function g as

* As before, f is an arbitrary bounded function.

a difference of two increasing functions for which the integrals $\int_a^b f dh_1$ and $\int_a^b f dh_2$ exist we have $g_1 + h_2 = g_2 + h_1$, and formula (28) shows that

$$\int_a^b f dg_1 + \int_a^b f dh_2 = \int_a^b f dg_2 + \int_a^b f dh_1$$

whence

$$\int_a^b f dg_1 - \int_a^b f dg_2 = \int_a^b f dh_1 - \int_a^b f dh_2$$

Let us prove that if the function f is continuous on the interval $[a, b]$ the integral $\int_a^b f dg$ is sure to exist. It is suffi-

cient to consider the case when g is an increasing function. By the uniform continuity of the function f , given an arbitrary $\varepsilon > 0$, there is $\delta > 0$ such that if the lengths $x_i - x_{i-1}$ of all the subintervals entering into the partition τ are less than δ we have $M_i - m_i < \varepsilon$ for all i . Then

$$\begin{aligned} S(\tau) - s(\tau) &= \sum_{i=1}^p (M_i - m_i) [g(x_i) - g(x_{i-1})] \leq \\ &\leq \varepsilon \sum_{i=1}^p [g(x_i) - g(x_{i-1})] = \varepsilon [g(b) - g(a)] \end{aligned}$$

whence follows the coincidence of the supremum and the infimum in inequality (27).

We shall not dwell in more detail on the elementary properties of the Stieltjes integral referring the reader to comprehensive courses in the theory of functions of a real variable (e.g. see I. P. Natanson, *op. cit.**). We shall confine

* In the book by I. P. Natanson as well as in many others the Stieltjes integral is defined in the classical way as the limit of the "Riemann-Stieltjes sums". The definition given in our course is only applicable when g is a function of bounded variation. In this case our definition is slightly more general than the classical one since it leads to a wider class of integrable functions. The definition equivalent

ourselves to establishing a relationship between the Stieltjes integral and the integral with respect to an additive set function. To this end we use Example 3, Sec. XI.1 and construct, as was done in this example, the additive set function φ , corresponding to the given increasing function g , defined on the algebra $\mathfrak{S}$ generated by the system of all cells $\Delta \subset [a, b)$.

Let us begin with the case when f is a continuous function on the interval $[a, b]$ and $E = [a, b)$. It follows from Sec. XI.3 that the integral $\int_E f d\varphi$ exists. On the other hand, every partition τ of the interval $[a, b]$ generates a partition of the cell E into the disjoint cells $e_i = [x_{i-1}, x_i)$ ($i = 1, 2, \dots, p$). Due to the continuity of the function f , its supremum and infimum on each subinterval $[x_{i-1}, x_i]$ ($i = 1, 2, \dots, p$) coincide, respectively, with those on the corresponding cell e_i , and therefore the Stieltjes-Darboux sums go into the Lebesgue-Darboux sums. For instance, for the upper sums we have

$$S(\tau) = \sum_{i=1}^p M_i [g(x_i) - g(x_{i-1})] = \sum_{i=1}^p M_i \varphi(e_i)$$

Hence,

$$\int_a^b f dg = \int_{[a, b)} f d\varphi \quad (29)$$

(here we have the Stieltjes integral over the interval $[a, b]$ on the left-hand side and the integral with respect to the additive set function φ over the cell $[a, b)$ on the right-hand side).

The same result can also be obtained when g is an arbitrary function of bounded variation on the interval $[a, b]$.

If the requirement that the function f should be continuous is rejected it can only be proved that the existence

(for functions g of bounded variation) to the one stated here can be found in the book by E. Kh. Gokhman, *The Stieltjes Integral and its Applications*, Moscow, Fizmatgiz, 1958 where the reader can also find some information on the relationship between this definition and the classical one.

of the Stieltjes integral $\int_a^b f dg$ implies the existence of the integral $\int_{[a, b)} f d\varphi$ and their coincidence. Indeed, in this case we have

$$M'_i = \sup_{x \in e_i} f(x) \leq M_i \text{ and } m'_i = \inf_{x \in e_i} f(x) \geq m_i$$

Therefore, assuming that the function g is increasing (it suffices to consider this case alone) we arrive at the following inequality for the Stieltjes-Darboux sums:

$$s(\tau) \leq \sum_{i=1}^p m'_i \varphi(e_i) \leq \sum_{i=1}^p M'_i \varphi(e_i) \leq S(\tau)$$

This means that between any given Stieltjes-Darboux sums there always lie some Lebesgue-Darboux sums. But $\sup_{\tau} s(\tau) = \inf_{\tau} S(\tau)$, and therefore the supremum and the infimum of the Lebesgue-Darboux sums also coincide, i.e. the integral $\int_{[a, b)} f d\varphi$ exists. It is also clear that equality (29) holds.

Let us consider an example showing that the converse is not true in this case. To this end we take the function

$$f(x) = g(x) = \begin{cases} 0 & \text{for } 0 \leq x < \frac{1}{2} \\ 1 & \text{for } \frac{1}{2} \leq x \leq 1 \end{cases}$$

and put $E = [0, 1)$. Let us split E into the two disjoint cells $e_1 = [0, \frac{1}{2})$ and $e_2 = [\frac{1}{2}, 1)$. Then $m'_1 = M'_1 = 0$, $m'_2 = M'_2 = 1$, $\varphi(e_1) = 1$, $\varphi(e_2) = 0$ and the Lebesgue-Darboux sums coincide. Consequently, the integral $\int_{[0, 1)} f d\varphi$ exists (and is equal to zero). At the same time the Stieltjes integral $\int_0^1 f dg$ does not exist. Indeed, for any partition τ of the interval $[0, 1]$ we can choose the index i so that $x_{i-1} <$

$< \frac{1}{2} \leq x_i$. Then $M_i = 1$, $m_i = 0$ and the difference $S(\tau) - s(\tau)$ turns out to be equal to 1:

$$S(\tau) - s(\tau) = g(x_i) - g(x_{i-1}) = 1$$

Thus, even in the classical case the passage from the integration in the Stieltjes sense to the integration with respect to the corresponding set function leads to a wider class of integrable functions.

§ 6. Distribution Functions on the Line

In this section we shall discuss in more detail the question mentioned in the last example in Sec. XI.1. We shall show that there is a correspondence between the measures defined on the semiring of cells in $\mathbf{R}_1$ and the increasing functions on the real line continuous from the left which makes it possible to reduce the integral with respect to a measure to the Stieltjes integral.

Definition. By a *distribution function* (on the line) we shall mean any finite wide-sense increasing (i.e. non-decreasing) function F continuous from the left and defined throughout the real axis $(-\infty, +\infty)$. Distribution functions differing from each other by a constant summand will be regarded as *equivalent*.

As was shown in Sec. XI.1, every distribution function generates (under the additional condition $\varphi \emptyset = 0$) the non-negative additive set function φ determined by formula (14). This function is defined on the semiring $\mathfrak{R}$ of one-dimensional cells with finite lengths*. Using the continuity from the left of the function F we shall prove that the set function φ is σ -additive. In other words, we shall prove the following theorem.

Theorem XI.6.1. *The additive set function φ generated on the semiring $\mathfrak{R}$ of cells in $\mathbf{R}_1$ by any distribution function F is a measure**.*

* It should be stressed that, in contrast to Chapter V, we now include into $\mathfrak{R}$ only finite intervals of the type of $[a, b)$.

** If we change formula (14) in the definition of the function φ by putting $\varphi(\Delta) = F(\beta - 0) - F(\alpha - 0)$ the set function φ turns out to be a measure even without the requirement that F should be continuous from the left.

The proof of the theorem is elaborated by analogy with the proof of the σ -additivity of the volume of a cell (see Sec. V.2). The finite additivity of the function φ was mentioned above and is obvious. Following the course of the proof of Lemma V.2.2 we readily show that if $\Delta_1, \Delta_2, \dots, \Delta_p$, are disjoint cells (belonging to $\mathfrak{R}$) and $\bigcup_{k=1}^p \Delta_k \subset \Delta$ (Δ is also a cell) then $\sum_{k=1}^p \varphi \Delta_k \leq \varphi \Delta$. By analogy with Lemma V.2.3 we then prove that if $\Delta_1, \Delta_2, \dots, \Delta_p$ are arbitrary cells (contained in $\mathfrak{R}$) and Δ is a cell contained in the union $\bigcup_{k=1}^p \Delta_k$ then $\varphi \Delta \leq \sum_{k=1}^p \varphi \Delta_k$. Lastly, the σ -additivity of the function φ is established by analogy with Lemma V.2.4. Let us dwell in more detail on this part of the proof.

Let a cell Δ be represented as $\Delta = \bigcup_{k=1}^{\infty} \Delta_k$ where the cells Δ_k are disjoint. The foregoing discussion indicates that

$$\sum_{k=1}^{\infty} \varphi \Delta_k \leq \varphi \Delta$$

and it only remains to derive the opposite inequality.

Let $\Delta = [a, b)$ and $\Delta_k = [a_k, b_k)$ ($k = 1, 2, \dots$) where $a_k < b_k$.^{*} Given an arbitrary $\varepsilon > 0$, by the continuity from the left of the function F , there are numbers $\delta_k > 0$ such that

$$F(a_k - \delta_k) > F(a_k) - \frac{\varepsilon}{2k}$$

Let us also set an arbitrary $\delta > 0$. The system of the intervals $(a_k - \delta_k, b_k)$ forms a covering of the closed interval $[a, b - \delta]$, and therefore it contains a finite subcovering of that interval. Consequently the cell $[a, b - \delta)$ can be covered by a finite system of cells $[a_k - \delta_k, b_k)$ ($k =$

^{*} It is sufficient to assume that all the cells Δ_k are nonempty.

$= 1, 2, \dots, p$). Hence, according to the above,

$$\begin{aligned} F(b - \delta) - F(a) &\leq \sum_{k=1}^p [F(b_k) - F(a_k - \delta_k)] < \\ &< \sum_{k=1}^p [F(b_k) - F(a_k)] + \varepsilon = \sum_{k=1}^p \varphi \Delta_k + \varepsilon \leq \sum_{k=1}^{\infty} \varphi \Delta_k + \varepsilon \end{aligned}$$

On passing to the limit as $\delta, \varepsilon \rightarrow 0$ we obtain

$$\varphi \Delta = F(b) - F(a) \leq \sum_{k=1}^{\infty} \varphi \Delta_k$$

which completes the proof.

An analogous theorem can be proved in the case when the definition of a distribution function includes the condition of the continuity from the right (instead of the continuity from the left). In this case the semiring of cells of the type of $[a, b)$ should be replaced by the semiring of intervals of the type of $(a, b]$.

Theorem XI.6.2. *Let ν be a finite measure on the semiring $\mathfrak{R}$ of cells in $\mathbf{R}_1$. Then the function F defined throughout the real line by the formula*

$$F(x) = \begin{cases} 0 & \text{for } x = 0 \\ \nu[0, x) & \text{for } x > 0 \\ -\nu[x, 0) & \text{for } x < 0 \end{cases}$$

is a distribution function, and the σ -additive set function φ generated by it coincides with ν . Any other distribution function serving as a generating function for the same measure ν is equivalent to F .

Proof. It is clear that for any cell $\Delta = [a, b)$ we have

$$F(b) - F(a) = \nu \Delta$$

Let us verify that the function F is continuous from the left (all the other properties of a distribution function are obvious).

For definiteness, let $x_0 > 0$ and let a (strictly) increasing sequence $\{x_n\}$ converge to x_0 : $x_n \rightarrow x_0 - 0$. We can assume that $x_n > 0$ for all n . Let us consider the cells $\Delta_n = [x_n, x_{n+1})$ and $\Delta = [x_1, x_0)$. Obviously, Δ_n 's are disjoint

and $\Delta = \bigcup_{n=1}^{\infty} \Delta_n$; therefore

$$F(x_0) - F(x_1) = \nu \Delta = \sum_{n=1}^{\infty} \nu \Delta_n = \sum_{n=1}^{\infty} [F(x_{n+1}) - F(x_n)]$$

whence $F(x_0) = \lim F(x_n)$.

If a distribution function G is also a generating function of the measure ν then

$$G(x) - G(0) = \nu[0, x) = F(x) \text{ for any } x > 0$$

and

$$G(x) - G(0) = -\nu[x, 0) = F(x) \text{ for any } x < 0$$

Consequently $G(x) = F(x) + G(0)$ for all x . The proof of the theorem has been completed.

A distribution function F can be discontinuous from the right. Let us elucidate the meaning of such discontinuities. For an arbitrary x_0 we take a decreasing sequence of the cells $\Delta_n = \left[x_0, x_0 + \frac{1}{n} \right)$. Then

$$F(x_0 + 0) - F(x_0) = \lim F\left(x_0 + \frac{1}{n}\right) - F(x_0) = \lim \varphi \Delta_n$$

where φ is the measure generated by the function F . But $\bigcap_{n=1}^{\infty} \Delta_n = (x_0)$, and hence $\lim \varphi \Delta_n$ is equal to the measure of the single-element set (x_0) which is attributed to this set in the extension of the measure φ from the semiring of cells to some σ -algebra. Thus, *the points of discontinuity of the function F are those to which strictly positive values of the measure are assigned.*

Let the reader check that for the extension of the measure φ (we retain the same symbol φ for the extended measure) we have, for any finite a and b ($a < b$), the relations

$$\varphi[a, b] = F(b + 0) - F(a)$$

$$\varphi(a, b) = F(b) - F(a + 0)$$

and

$$\varphi(a, b] = F(b + 0) - F(a + 0)$$

Furthermore, $\varphi(-\infty, +\infty) = \lim_{x \rightarrow +\infty} F(x) - \lim_{x \rightarrow -\infty} F(x)$.

Let us demonstrate the way in which the integral with respect to measure (in an abstract space) can be reduced to the Stieltjes integral. We shall limit ourselves to the case when the integrand function is bounded above since the passage to unbounded functions would involve, instead of the Stieltjes integral over a finite interval, the integral of the same type taken over the entire real line.

Thus, let X be an abstract measure space with measure μ defined on a σ -algebra and f be a bounded measurable function defined on a set $E \subset X$ with $\mu E < +\infty$; also let $A < f(x) < B$ for all $x \in E$. Putting

$$F(t) = \mu E [f(x) < t] \quad (30)$$

we obtain a function defined for all t and assuming nonnegative values. Let us verify that F is a distribution function on the line. The monotonicity of F is evident; the continuity from the left is proved as follows. If an increasing sequence t_n tends to t_0 ($t_n \rightarrow t_0 - 0$) we have

$$E [f(x) < t_0] = \bigcup_{n=1}^{\infty} E [f(x) < t_n]$$

and, by Theorem IV.1.2, $F(t_n) \rightarrow F(t_0)$. Moreover, $F(t) = 0$ for $t \leq A$ and $F(t) = \mu E$ for $t \geq B$.

Let us prove the equality

$$\int_E f d\mu = \int_A^B t dF(t) \quad (31)$$

where the left-hand member is the Lebesgue integral and the right-hand member is the Stieltjes integral (whose existence is guaranteed by the continuity of the integrand function $g(t) = t$).

To this end, we take an arbitrary partition τ of the interval $[A, B]$ with the aid of points of division $A = t_0 < t_1 < \dots < t_p = B$ and put

$$e_i = E [t_{i-1} \leq f(x) < t_i] \quad (i = 1, 2, \dots, p)$$

Then $\mu e_i = F(t_i) - F(t_{i-1})$, and the Stieltjes-Darboux sums corresponding to the function $g(t) = t$ are written as

$$S(\tau) = \sum_{i=1}^p t_i [F(t_i) - F(t_{i-1})] = \sum_{i=1}^p t_i \mu e_i$$

and

$$s(\tau) = \sum_{i=1}^p t_{i-1} [F(t_i) - F(t_{i-1})] = \sum_{i=1}^p t_{i-1} \mu e_i$$

The partition τ of $[A, B]$ generates the partition τ' of the set E into disjoint sets e_i . Putting, as usual,

$$m_i = \inf_{x \in e_i} f(x) \text{ and } M_i = \sup_{x \in e_i} f(x)$$

we obtain $t_{i-1} \leq m_i \leq M_i \leq t_i$ for any i . Consequently, the Lebesgue-Darboux sums $s(\tau'; f)$ and $S(\tau'; f)$ lie between $s(\tau)$ and $S(\tau)$:

$$s(\tau) \leq s(\tau'; f) \leq S(\tau'; f) \leq S(\tau)$$

It follows that

$$s(\tau) \leq \int_E f d\mu \leq S(\tau)$$

for any τ . But we have

$$\int_A^B t dF(t) = \sup_{\tau} s(\tau) = \inf_{\tau} S(\tau)$$

and equality (31) has thus been proved.

The material presented in this section has important applications in the *theory of probability* (where the notion of a distribution function originally appeared). Let x be a *random variable* assuming finite real values. The function F defined for all $t \in (-\infty, +\infty)$ as the probability μ of the fulfilment of the inequality $x \leq t$ is called the *probability distribution of the random variable x* . It is supposed that the probability μ is defined on some σ -algebra $\mathfrak{S}$ in $\mathbf{R}_1$ containing all the Borel sets; it is a nonnegative σ -additive set function, that is a measure, normalized by the condition $\mu(-\infty, +\infty) = 1$ which means that the probability of x taking some real value is equal to unity. Therefore the distribution functions considered in the theory of probability increase from 0 to 1 along the real line.

Let f be a bounded $\mathfrak{S}$ -measurable (i.e. measurable with respect to the σ -algebra $\mathfrak{S}$) function defined on the real axis; then $A < f(x) < B$ for all $x \in \mathbf{R}_1$. The quantity

$y = f(x)$ is also a random variable and its probability distribution coincides with the function F specified by formula (30). The integral $\int_A^B t dF(t)$ expresses the so-called *mathematical expectation* of the random variable y . We thus see that the integral $\int_{-\infty}^{+\infty} f d\mu$ is nothing but another representation of this same mathematical expectation.

We shall demonstrate what has been said by the example of a *discrete random variable*. Let the probability μ be concentrated on certain single-element sets $(t_1), (t_2), \dots$ and let the values of μ at these points be, respectively, nonzero numbers $\mu_1, \mu_2, \dots$. Then we must have $\mu_i > 0$ and $\sum_{i=1}^{\infty} \mu_i = 1$. In other words, the random variable x in question can only assume one of the values t_i , the number μ_i being the probability of the equality $x = t_i$. Then $y = f(x)$ is also a discrete random variable only assuming one of the values $f(t_i)$, and the mathematical expectation of y is equal to

$$\sum_{i=1}^{\infty} f(t_i) \mu_i$$

This sum is nothing but the integral $\int_{-\infty}^{+\infty} f d\mu$ (cf. Example 1 in Sec. VII.1).

Absolutely Continuous Set Functions

§ 1. Definition of an Absolutely Continuous Set Function

In the foregoing chapters, when studying the Lebesgue integral, we proved its absolute continuity (Theorem VIII.3.3). In this chapter we shall present a general investigation of σ -additive set functions possessing the property of absolute continuity and show that this property completely characterizes the functions representable as integrals. But the definition of absolute continuity will be stated here in another form; the equivalence of the new and the old definitions will be proved for functions assuming finite values. When considering additive functions which can take infinite values we shall suppose that the value $-\infty$ is excluded.

Definition. Let X be a fixed measure space with a σ -finite measure μ defined on a σ -algebra $\mathfrak{S}$ and φ be an arbitrary σ -additive set function defined on the same σ -algebra $\mathfrak{S}$.* The function φ is said to be *absolutely continuous with respect to the measure μ* (in this case we shall write $\varphi \ll \mu$) if $\varphi(A) = 0$ for any set $A \in \mathfrak{S}$ of μ -measure zero ($\mu A = 0$).

From the definition it follows directly that if $\varphi \ll \mu$ then all the three variations of φ are also absolutely continuous with respect to μ . Conversely, if we have $|\varphi| \ll \mu$ for the total variation $|\varphi|$ of φ then $\varphi \ll \mu$.

Theorem XII.1.1. *For a σ -additive set function φ with finite values defined on the σ -algebra $\mathfrak{S}$ to be absolutely continuous with respect to the measure μ it is necessary and sufficient that for any $\varepsilon > 0$ there should exist $\delta > 0$ such that $|\varphi(A)| < \varepsilon$ for every $A \in \mathfrak{S}$ with $\mu A < \delta$.*

The last condition can be written in the shortened form as

$$\lim_{\mu A \rightarrow 0} \varphi(A) = 0 \quad (1)$$

* A σ -additive set function defined on a σ -algebra and in general taking values of both signs is also called a *signed measure* or *charge*.—Tr.

Proof. For any σ -additive set function φ the sufficiency of condition (1) is evident. Let us prove its necessity.

Suppose that, despite the absolute continuity of the function φ , condition (1) is violated. By Theorem XI.1.4, the function φ is bounded. Let us arbitrarily set positive

numbers δ_n so that $\sum_{n=1}^{\infty} \delta_n < +\infty$. Then there is a certain

$\varepsilon > 0$ for which there are some sets $A_n \in \mathfrak{S}$ with $\mu A_n < \delta_n$ ($n = 1, 2, \dots$) such that $|\varphi(A_n)| \geq \varepsilon$, and, moreover, $|\varphi|(A_n) \geq \varepsilon$. Let us put

$$B_p = \bigcup_{n=p}^{\infty} A_n \text{ and } B = \bigcap_{p=1}^{\infty} B_p$$

Then $\mu B = 0$, and consequently $|\varphi|(B) = 0$. But $|\varphi|(B_p) \geq |\varphi|(A_p) \geq \varepsilon$ for any p , which contradicts Theorem IV.1.2.

Let us consider an example showing that the theorem proved no longer holds if the condition that the values of the function φ are finite is violated. The function φ considered in this example is σ -finite. Let us break up the space R_1 into a countable number of disjoint sets E_n with $\mu E_n = 1$ (here μ is the Lebesgue measure) and put

$$\varphi(A) = \sum_{n=1}^{\infty} n \mu(A \cap E_n)$$

for every measurable set $A \subset R_1$. The function φ is clearly continuous but condition (1) is not fulfilled for it since we have $\varphi(A) = n \cdot \mu A$ for any measurable set $A \subset E_n$ and hence $\varphi(A)$ can be arbitrarily large when μA is arbitrarily small.

The next theorem indicates that condition (1) characterizes not only the absolute continuity but, under certain conditions, also the σ -additivity of set functions.

Theorem XII.1.2. *If a finitely additive function φ defined on a σ -algebra satisfies condition (1) and the measure μ is finite the function φ is σ -additive.*

Proof. Let $A = \bigcup_{n=1}^{\infty} A_n$ where the sets $A_n \in \mathfrak{S}$ are disjoint. Putting $B_p = \bigcup_{n=1}^p A_n$ we have $\mu(A \setminus B_p) \rightarrow 0$ and,

consequently, by condition (1), there must be $\varphi(A \setminus B_p) \rightarrow 0$, which means that

$$\varphi(A) = \sum_{n=1}^{\infty} \varphi(A_n)$$

§ 2. The Radon-Nikodym Theorem

We now pass to the fundamental theorem of the present chapter. As before, we suppose that X is a measure space with a σ -finite measure μ defined on a σ -algebra $\mathfrak{S}$.

Theorem XII.2.1. (J. Radon-O. Nikodym*). *If φ is a finite σ -additive set function, defined on $\mathfrak{S}$, assuming finite values and absolutely continuous with respect to μ , there exists a μ -summable function f on X such that*

$$\varphi(A) = \int_A f d\mu \quad (2)$$

for any $A \in \mathfrak{S}$.

The function f in (2) is determined uniquely to within the equivalence of functions, that is if g is another μ -summable function satisfying the same condition (2) there must be $f \sim g$.

Proof. Since $\varphi = \varphi_+ - \varphi_-$ and φ_+, φ_- are also σ -additive functions with finite values (see Theorems XI.1.3 and XI.1.4) absolutely continuous with respect to μ , it is sufficient to elaborate the proof for the case when φ is nonnegative, that is when φ is a measure in X .

We shall begin with the assumption that the measure μ also takes only finite values. Let us put $\omega = \varphi + \mu$. This is also a measure in X defined on $\mathfrak{S}$ and taking finite values. Consider the space $L^2_X(\omega)$ constructed on X with measure ω , that is consisting of the functions square summable on X with respect to the measure ω . For any measurable function g we have

$$\int_X g^2 d\varphi \leq \int_X g^2 d\omega$$

(see the remark after Lemma XI.2.1), and therefore if $g \in L^2_X(\omega)$ then $g \in L^2_X(\varphi)$. Now, since $\varphi(X) < +\infty$, we conclude, by Theorem IX.1.1, that g is φ -summable. Conse-

* O. Nikodym, a 20th century Polish mathematician.

quently, the integral

$$F(g) = \int_X g d\varphi$$

represents a distributive functional defined on $L^2_X(\omega)$. Let us show that $F(g)$ is a linear functional. According to Theorem III.7.2, it is sufficient to verify that it is bounded. To this end we use Bunyakovsky's inequality:

$$|F(g)|^2 \leq \int_X g^2 d\varphi \int_X 1^2 d\varphi \leq \varphi(X) \int_X g^2 d\omega = \varphi(X) \|g\|^2$$

By Theorem IX.4.1 on the general representation of a linear functional in the space L^2 , there is a function $h \in L^2_X(\omega)$ such that

$$F(g) = \int_X gh d\omega$$

for any $g \in L^2_X(\omega)$. Whence

$$\int_X g d\varphi = \int_X gh d\omega \quad (3)$$

In particular, putting g equal to the characteristic function of an arbitrary set $A \in \mathfrak{S}$ we see that

$$\varphi(A) = \int_A h d\omega \quad (4)$$

It follows that the inequality $h(x) \geq 0$ is fulfilled ω -almost everywhere on X . Indeed, if we suppose that $h(x) < 0$ on a set A_0 with $\omega A_0 > 0$ formula (4) and Theorem VIII.3.6 show that

$$\varphi(A_0) = \int_{A_0} h d\omega < 0$$

while, by the hypothesis, the function φ is nonnegative. Furthermore, we have $\varphi(A) \leq \omega(A)$ for any $A \in \mathfrak{S}$ and therefore

$$\int_A h d\omega \leq \omega(A)$$

whence, by the same argument, the inequality $h(x) \leq 1$ holds ω -almost everywhere on X . Without losing generality we can now assume that $0 \leq h(x) \leq 1$ for all $x \in X$.

Let us take the set

$$N = X [h(x) = 1]$$

With the aid of formula (4) we obtain

$$\varphi(N) = \int_N h d\omega = \omega(N) = \varphi(N) + \mu N$$

and consequently $\mu N = 0$, that is $h(x) < 1$ throughout the set X except possibly at the points of the set N with $\mu N = 0$. It should be noted that the absolute continuity of the function φ shows that we also have $\varphi(N) = 0$.

From equality (3) it follows that

$$\int_X g(1-h) d\varphi = \int_X gh d\omega - \int_X gh d\varphi$$

for any $g \in L_X^2(\omega)$ and hence, by Lemma XI.2.1,

$$\int_X g(1-h) d\varphi = \int_X gh d\mu \quad (5)$$

In the general case when g is an arbitrary nonnegative measurable everywhere finite function on X the corresponding truncated functions g_m belong to $L_X^2(\omega)$. On writing down equality (5) for g_m and applying Levi's theorem we conclude that the same equality (5) also holds for g .

Now we define the sought-for function f by putting

$$f(x) = \begin{cases} \frac{h(x)}{1-h(x)} & \text{for } x \in X \setminus N \\ 0 & \text{for } x \in N \end{cases}$$

Let us check that equality (2) holds. To this end we take the function g specified by the formula

$$g(x) = \begin{cases} \frac{\chi_A(x)}{1-h(x)} & \text{for } x \in X \setminus N \\ 0 & \text{for } x \in N \end{cases}$$

where χ_A is the characteristic function of the set A . This function g is obviously nonnegative, measurable and has

finite values. Substituting it into equality (5) we see that

$$\varphi(A) = \int_X gh \, d\mu = \int_A \frac{h}{1-h} d\mu = \int_A f \, d\mu$$

for any arbitrary set $A \in \mathfrak{S}$.^{*} Thus, the existence of the desired function f has been proved and the constructed function f is nonnegative.

Now we pass to the case when the measure μ is σ -finite. Let us break up the space X into a countable number of disjoint subsets X_i with $\mu X_i < +\infty$ ($i = 1, 2, \dots$). By what has been proved, for each of the sets X_i there exists a nonnegative summable function f_i such that

$$\varphi(A) = \int_A f_i \, d\mu$$

for any measurable set $A \subset X_i$. To define the required function f on X we put $f(x) = f_i(x)$ for $x \in X_i$ ($i = 1, 2, \dots$). Then f is measurable (see proposition 3° in Sec. VI.1), nonnegative and finite almost everywhere on X . On the one hand, for any $A \in \mathfrak{S}$ we have

$$\varphi(A) = \sum_{i=1}^{\infty} \varphi(A_i) \text{ where } A_i = A \cap X_i$$

and, on the other hand, by Lemma VIII.2.1,

$$\int_A f \, d\mu = \sum_{i=1}^{\infty} \int_{A_i} f \, d\mu = \sum_{i=1}^{\infty} \int_{A_i} f_i \, d\mu = \sum_{i=1}^{\infty} \varphi(A_i)$$

These relations prove equality (2). In particular, putting $A = X$ in (2) we see that f is summable on X .

Let us prove the uniqueness of the function f . Suppose that g is another function satisfying equality (2). Then the difference $l = f - g$ satisfies the relation

$$\int_A l \, d\mu = 0$$

for any $A \in \mathfrak{S}$.[†]

^{*} Equality $\mu N = \varphi(N) = 0$ allows us to discard the set N in the integration process.

If we suppose that there is a set of positive measure at whose all points either $l(x) > 0$ or $l(x) < 0$ we arrive at a contradiction. Consequently $l(x) = 0$ almost everywhere, i.e. $f \sim g$.

As was noted in the course of the proof, the function f is nonnegative. Thus, if the function φ is nonnegative any function f satisfying condition (2) of the Radon-Nikodym theorem is such that $f(x) \geq 0$ almost everywhere on X .

Now let us extend the Radon-Nikodym theorem to the case when the given function φ is only σ -finite (and satisfies all the other conditions of the theorem). By the corollary of Theorem XI.1.4, the function φ_+ is then also σ -finite. Let us represent the space X as a countable union of disjoint subsets X_i ($i = 1, 2, \dots$) on each of which $\varphi_+(X_i)$ is finite. By the above, for every subset X_i there exists a nonnegative summable function g_i such that

$$\varphi_+(A) = \int_A g_i d\mu$$

for any measurable set $A \subset X_i$.

Putting

$$g(x) = g_i(x) \quad \text{for } x \in X_i \quad (i = 1, 2, \dots)$$

and arguing by analogy with the end of the proof of Theorem XII.2.1 we readily show that

$$\varphi_+(A) = \int_A g d\mu$$

for any $A \in \mathfrak{S}$.

Since the function φ does not assume the value $-\infty$ Theorem XI.1.5 implies that it is bounded below and therefore φ_- is finite. Now we can use Theorem XII.2.1 according to which there exists a function h summable on X such that

$$\varphi_-(A) = \int_A h d\mu \quad \text{for any } A \in \mathfrak{S}$$

It now becomes clear that $f = g - h$ is the sought-for function, that is it satisfies condition (2) for any $A \in \mathfrak{S}$. But, in contrast to the case of a finite function φ , the function f constructed here is not necessarily summable throughout the space X .

The function f determined from a given function φ on the basis of the Radon-Nikodym theorem is called *the Radon-Nikodym derivative of φ with respect to μ* (or *the density of the charge φ*) and is sometimes denoted as $\frac{d\varphi}{d\mu}$. It should be noted that the Radon-Nikodym theorem establishes the existence of the derivative $\frac{d\varphi}{d\mu}$ without telling how to find its values at given points and only determines it "globally" (to within the equivalence of functions) by means of relation (2). Some indications concerning the possibility of the determination of concrete values of the Radon-Nikodym derivative will be given in the next chapter.

The Radon-Nikodym derivative possesses a property analogous to that of an ordinary derivative connected with change of variable under the integral sign. This becomes clear if we rewrite equality (19) of Chapter XI in the form

$$\int_E f d\varphi = \int_E f \frac{d\varphi}{d\mu} d\mu$$

§ 3. The Hahn Decomposition

The Radon-Nikodym theorem enables us to establish an interesting property of charges (i.e. of σ -additive set functions) known as *the Hahn decomposition*.*

Theorem XII.3.1. *Let φ be a σ -finite and σ -additive function defined on a σ -algebra $\mathfrak{S}$ of subsets of a set X . Then there are two sets $X_1, X_2 \in \mathfrak{S}$ such that*

- (a) $X_1 \cup X_2 = X$;
- (b) $X_1 \cap X_2 = \emptyset$;
- (c) $\varphi_-(X_1) = \varphi_+(X_2) = 0$.

The last property means that $\varphi(E) \geq 0$ for all the subsets $E \subset X_1$ and $\varphi(E) \leq 0$ for all the subsets $E \subset X_2$ ($E \in \mathfrak{S}$).

Proof. The total variation $|\varphi|$ is a σ -finite measure in X (see the Corollary of Theorem XI.1.4). Let us denote it as μ . It follows from the definition of the total variation

* H. Hahn (1879-1934), an Austrian mathematician. A direct (but more complicated) derivation of the Hahn decomposition can be, for instance, found in the book by P. R. Halmos, *Measure Theory*, van Nostrand, Princeton, 1950.

that $\varphi \ll \mu$. Consequently, there exists the Radon-Nikodym derivative $f = \frac{d\varphi}{d\mu}$. Now we put

$$X_1 = X [f(x) \geq 0] \text{ and } X_2 = X [f(x) < 0]$$

It is clear that $X_1, X_2 \in \mathfrak{S}$ and that these sets satisfy conditions (a) and (b), and condition (c) is implied by formula (13) of Chapter XI.

Now we can give the Radon-Nikodym theorem a further generalization. Let there be given two σ -finite and σ -additive set functions φ and ω defined on a σ -algebra $\mathfrak{S}$ of subsets of a set X . Let us agree that φ will be called absolutely continuous with respect to ω if $\varphi \ll |\omega|$.

Let X_1 and X_2 be two sets satisfying, for the function ω , all the conditions of the above theorem. Then

$$|\omega|(E) = \omega(E) \text{ if } E \in \mathfrak{S} \text{ and } E \subset X_1$$

and

$$|\omega|(E) = -\omega(E) \text{ if } E \in \mathfrak{S} \text{ and } E \subset X_2$$

Finally, let φ be absolutely continuous with respect to ω and $g = \frac{d\varphi}{d|\omega|}$. Then, putting

$$f(x) = \begin{cases} g(x) & \text{for } x \in X_1 \\ -g(x) & \text{for } x \in X_2 \end{cases}$$

we obtain, for any $A \in \mathfrak{S}$, the relation

$$\begin{aligned} \varphi(A) &= \int_A g d|\omega| = \int_{A \cap X_1} f d\omega + \int_{A \cap X_2} (-f) d(-\omega) = \\ &= \int_{A \cap X_1} f d\omega + \int_{A \cap X_2} f d\omega = \int_A f d\omega \end{aligned}$$

Indefinite Lebesgue Integral

§ 1. Absolutely Continuous Point Functions

In this chapter we shall dwell in more detail on the integral of a function of one variable, that is on the integral in the one-dimensional space R_1 . The integrand function f will be supposed to be defined and summable on an interval $[a, b]$. Then for any $x \in [a, b]$ there exists the function

$$F(x) = \int_a^x f dx$$

that is the integration process directly leads not only to the interpretation of the integral as a set function but also to the concept of the integral "with variable upper limit" which is a point function. In this connection the following questions naturally arise: what is the class of functions F of one variable x representable as the integral of a summable function f with variable limit of integration and is it possible, as in the case of the classical (Riemann) integral of a continuous function, to reconstruct the integrand f from the given function F by means of differentiation? To answer these questions we need the following

Definition. A function F (with finite values) defined on an interval $[a, b]$ is said to be *absolutely continuous* on that interval if, given any $\varepsilon > 0$, there is $\delta > 0$ such that for any finite system of intervals $[a_i, b_i]$ ($i = 1, 2, \dots, n$) having no interior points in common* and contained in $[a, b]$ whose sum of lengths is less than δ (i.e. $\sum_{i=1}^n (b_i - a_i) < \delta$)

* Thus, here we deal with *nonoverlapping* intervals (cf. Sec. V.1), and the case when some of the intervals $[a_i, b_i]$ have common end points is not excluded.

$< \delta$) there holds the inequality

$$\sum_{i=1}^n |F(b_i) - F(a_i)| < \varepsilon \quad (1)$$

It is clear that the absolute continuity of the function F implies its uniform continuity (which is readily seen if we take a system consisting of only one interval) and hence its continuity at every point of the interval $[a, b]$.

Remark. Inequality (1) in the definition of absolute continuity can be replaced by a weaker condition

$$\left| \sum_{i=1}^n [F(b_i) - F(a_i)] \right| < \varepsilon \quad (2)$$

Indeed, given an arbitrary $\varepsilon > 0$, let $\delta > 0$ be such that inequality (2) is fulfilled for any finite system of nonoverlapping subintervals whose sum of lengths is less than δ . Taking any such system of intervals $\{[a_i, b_i]\}$ ($i = 1, 2, \dots, n$) we can split the intervals entering into this system into two groups the first of which includes those intervals for which $F(b_i) - F(a_i) \geq 0$ while the other includes all the remaining intervals. By the hypothesis, inequality (2) must be fulfilled for each of the groups separately and therefore the sum of the *absolute values* of the increments of the function F gained on the intervals belonging to each group is less than ε whence

$$\sum_{i=1}^n |F(b_i) - F(a_i)| < 2\varepsilon$$

By the arbitrariness of ε , this means that the function F is absolutely continuous.

A simple example of an absolutely continuous function defined on an interval $[a, b]$ is a function F satisfying the *Lipschitz* condition* requiring that there should exist a constant C such that

$$|F(x_2) - F(x_1)| \leq C |x_2 - x_1|$$

for any x_1 and x_2 belonging to that interval. In particular, every function differentiable at each point of the interval $[a, b]$ is absolutely continuous.

* R. O. S. Lipschitz (1832-1903), a German mathematician.

It is clear that the sum and the difference of two absolutely continuous functions are also absolutely continuous.

Let us agree that throughout this chapter the letter μ will denote the Lebesgue measure on the line.

Now, by analogy with Example 3 in Sec. XI.1, we shall assign to every function F absolutely continuous on an interval $[a, b]$ a certain additive set function. Namely, let $\mathfrak{S}$ be the algebra of subsets of the cell $I = [a, b)$ generated

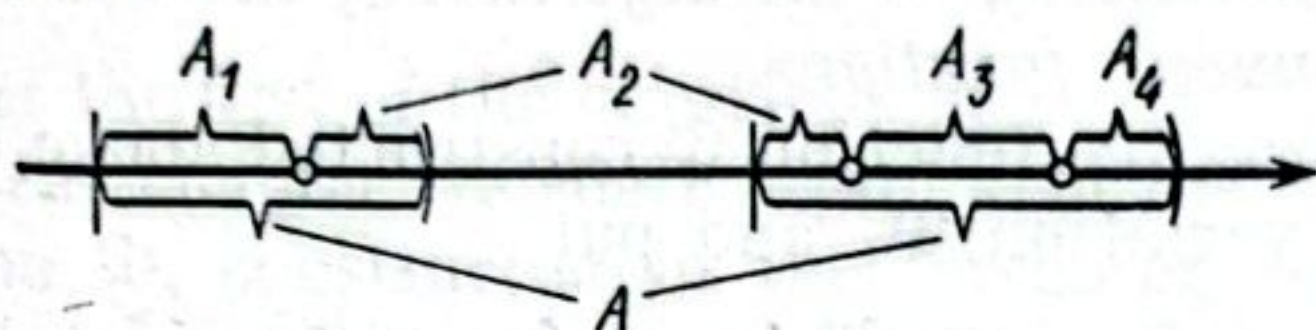


Fig. 20

by the system of all cells contained in I . If A is a set belonging to $\mathfrak{S}$ and $A = \bigcup_{i=1}^p \Delta_i$ is its representation as a finite union of disjoint cells $\Delta_i = [a_i, b_i)$ we put

$$\Phi(A) = \sum_{i=1}^p [F(b_i) - F(a_i)] \quad (3)$$

It readily follows from the definition of the absolute continuity of the function F that $\lim_{\mu A \rightarrow 0} \Phi(A) = 0$. More precisely, if $\mu A < \delta$ then $|\Phi(A)| < \varepsilon$ for any pair of numbers ε, δ mentioned in the definition of the absolute continuity of F .

Let us show that *the function Φ is bounded*. To this end, we take a pair of numbers ε, δ mentioned above. Any $A \in \mathfrak{S}$ can be represented in the form of a finite union of disjoint sets $A_j \in \mathfrak{S}$ ($j = 1, 2, \dots, l$) so that $\mu A_j = \frac{\delta}{2}$ for every $j \leq l-1$ and $\mu A_l \leq \frac{\delta}{2}$ (see Fig. 20). Then $(l-1) \frac{\delta}{2} \leq b-a$ whence $l \leq C$ where $C = \frac{2}{\delta}(b-a) + 1$. Since $|\Phi(A_j)| < \varepsilon$ for every $j = 1, 2, \dots, l$ there must be

$$|\Phi(A)| \leq \sum_{j=1}^l |\Phi(A_j)| < l\varepsilon \leq C\varepsilon$$

The same argument shows that *every absolutely continuous function of one variable is a function of bounded variation*. This also follows from the theorem below. By the way, it should be noted that the converse does not hold even if we confine ourselves to continuous functions, that is a continuous function of bounded variation may not be absolutely continuous.

Theorem XIII.1.1. *Every absolutely continuous point function F is representable as the difference of two increasing absolutely continuous functions.*

Proof. Using the function Φ constructed above, let us take its positive variation Φ_+ and put

$$G(x) = \begin{cases} \Phi_+[a, x) & \text{for } a < x \leq b \\ 0 & \text{for } x = a \end{cases}$$

It is obvious that G is an increasing function. Let us prove that it is absolutely continuous. Setting an arbitrary $\varepsilon > 0$ we take $\delta > 0$ mentioned in the definition of the absolute continuity of the function F and then consider an arbitrary system of nonoverlapping intervals $[a_i, b_i] \subset [a, b]$ ($i =$

$= 1, 2, \dots, n$) such that $\sum_{i=1}^n (b_i - a_i) < \delta$. Let us put

$A = \bigcup_{i=1}^n [a_i, b_i)$. Then $A \in \mathfrak{S}$ and $\mu A < \delta$. According to

the definition of the positive variation, we have

$$\Phi_+(A) = \sup_{B \subset A, B \in \mathfrak{S}} \Phi(B)$$

But $\mu B \leq \mu A < \delta$ and therefore $|\Phi(B)| < \varepsilon$, which means that $\Phi_+(A) \leq \varepsilon$. In other words,

$$\sum_{i=1}^n [G(b_i) - G(a_i)] \leq \varepsilon$$

which proves the absolute continuity of the function G .

Similarly, using the negative variation Φ_- we construct the increasing absolutely continuous function of one variable

$$H(x) = \begin{cases} \Phi_-[a, x) & \text{for } a < x \leq b \\ 0 & \text{for } x = a \end{cases}$$

Now, for any $x \in [a, b]$, we can write

$$F(x) = F(a) + \Phi[a, x) = F(a) + \Phi_+[a, x) - \Phi_-[a, x) = [F(a) + G(x)] - H(x)$$

which gives the required representation.*

§ 2. Properties of Functions Representable in the Form of an Integral

To answer the first of the questions posed at the beginning of the present chapter we shall begin with the extension of the function Φ , constructed in the foregoing section for a given absolutely continuous point function F , to a σ -algebra of sets. Theorem XIII.1.1 makes it possible to reduce this auxiliary problem to the case when the function F is increasing and, consequently, the function Φ is positive.

Thus, let F be an increasing absolutely continuous function defined on an interval $[a, b]$ and Φ be the set function defined by formula (3) on the algebra $\mathfrak{S}$ generated by the system of all the cells contained in $I = [a, b]$. We shall temporarily restrict the function Φ and consider it as being only defined on the semiring of cells contained in I . Theorem XI.6.1 indicates that Φ is a measure.** Let Ψ be its standard extension to a σ -algebra $\mathfrak{T}$ of some subsets of I . It is clear that $\mathfrak{S} \subset \mathfrak{T}$ and $\Phi(A) = \Psi(A)$ for any $A \in \mathfrak{S}$.

Theorem XIII.2.1. *The σ -algebra $\mathfrak{T}$ contains all the Lebesgue measurable subsets of I . If A is measurable and $\mu A = 0$ then $\Psi(A) = 0$.*

Proof. Let Ψ^* be the outer measure generated by the measure Φ . We shall show that if $A \subset I$ and $\mu A = 0$ then $\Psi^*(A) = 0$. To this end we set an arbitrary $\varepsilon > 0$ and take the corresponding number $\delta > 0$ mentioned in the definition of the absolute continuity of the function F . Then we find an open set G such that $A \subset G$ and $\mu G < \delta$. Let us represent G as the union of its components (a_i, b_i) ($i = 1, 2, \dots$).

* Since G is an increasing absolutely continuous function so is the sum $F(a) + G(x)$. If $x = a$ the cell $[a, a)$ is empty.

** To apply formally Theorem XI.6.1 we should extend the function F to the entire x -axis by putting, for example, $F(x) = F(a)$ for $x < a$ and $F(x) = F(b)$ for $x > b$. Then F becomes a distribution function.

Then

$$A \subset \bigcup_i [a_i, b_i)$$

where all the cells involved are disjoint. Without loss of generality we can assume that all the cells $[a_i, b_i)$ belong to I (since the set A itself belongs to I). By the definition of the outer measure,

$$\Psi^*(A) \leq \sum_i \Phi[a_i, b_i) = \sum_i [F(b_i) - F(a_i)]$$

If the number of summands on the right-hand side is finite, the whole sum is less than ε since $\sum_i (b_i - a_i) < \delta$. If the number of summands is infinite we have

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n [F(b_i) - F(a_i)] \leq \varepsilon$$

Thus, $\Psi^*(A) \leq \varepsilon$, which implies, by the arbitrariness of ε , that $\Psi^*(A) = 0$.

As is known, if $\Psi^*(A) = 0$ the set A is Ψ^* -measurable (see Sec. IV.3, proposition 1°) and, consequently, $A \in \mathfrak{L}^*$ and $\Psi(A) = 0$.

Now, let A be an arbitrary Lebesgue measurable subset of I . By Theorem V.5.6, we have $A = H \cup B$ where H is an F_δ -set, $B \cap H = \emptyset$ and $\mu B = 0$. By what has already been proved, $B \in \mathfrak{L}$. On the other hand, all the Borel subsets of I must be contained in $\mathfrak{L}$ since the σ -algebra of the Borel sets is the minimal σ -algebra containing all the cells. Consequently, $H \in \mathfrak{L}$ and therefore $A \in \mathfrak{L}$. The theorem has been proved.

Let $\mathfrak{M}$ denote the σ -algebra of all Lebesgue measurable subsets of the cell I . By Theorem XIII.2.1, $\mathfrak{M} \subset \mathfrak{L}$. It should be noted that $\mathfrak{L}$ can be essentially wider than $\mathfrak{M}$. For instance, if $F(x) = \text{const}$ on a subinterval $[\alpha, \beta] \subset [a, b]$, all the subsets of the cell $[\alpha, \beta)$ belong to $\mathfrak{L}$. In what follows we shall consider the restriction of the function Ψ to the σ -algebra $\mathfrak{M}$ retaining the same letter Ψ to denote it. Then the foregoing theorem implies that $\Psi \ll \mu$.

* We remind the reader that, according to the definition of the standard extension, the σ -algebra $\mathfrak{L}$ consists of all Ψ^* -measurable sets.

Theorem XIII.2.2. *For a function F of one variable defined on an interval $[a, b]$ to be representable in the form*

$$F(x) = F(a) + \int_a^x f dx \quad (4)$$

where f is a summable function on $[a, b]$ it is necessary and sufficient that F be absolutely continuous.

Proof. The necessity of the condition follows from the absolute continuity of the integral of a summable function (see the remark made in connection with inequality (2)). Let us prove its sufficiency. By virtue of Theorem XIII.1.1, we should only elaborate the proof for the case when F is an increasing function.

Let us use the above constructed σ -additive function Ψ on the σ -algebra $\mathfrak{M}$. Since it is absolutely continuous with respect to μ the Radon-Nikodym theorem guarantees its representability in the form of the integral of a summable function:

$$\Psi(A) = \int_A f dx \quad (A \in \mathfrak{M})$$

In particular, for any $x \in [a, b]$ we have

$$F(x) - F(a) = \Phi[a, x) = \Psi[a, x) = \int_a^x f dx$$

§ 3. Differentiation of Continuous Monotone Functions

As is known, a continuous function is not necessarily differentiable, and there are examples of continuous functions having no derivative at any point. But if we additionally suppose that the continuous function in question is monotone it turns out that it possesses a derivative everywhere except possibly at points forming a set of measure zero. Here we shall present the corresponding investigation for the case of an increasing function following the scheme suggested in the book by F. Riesz and Sz. Nagy, *Leçons d'analyse fonctionnelle*, Budapest, 1953 (see the bibliography).

Definition. Let g be a continuous function defined on an interval $[a, b]$. Then an interior point x of this interval is

said to be *invisible from the right* (with respect to g) if there is a point ξ_1 such that $x < \xi_1 \leq b$ and $g(\xi_1) > g(x)$ and *invisible from the left* if there is a point ξ_2 such that $a \leq \xi_2 < x$ and $g(\xi_2) > g(x)$. We shall briefly call the points of the first type *r-points* of the function $g(x)$ and denote the set of all such points by K_r or, more precisely, $K_r(g)$. Accordingly, for the point of the second type we shall use the term *l-points* and denote the set of these points K_l or $K_l(g)$.

It is evident that if $x_0 \in (a, b)$ is a point at which the function g attains its greatest value then x_0 is not contained

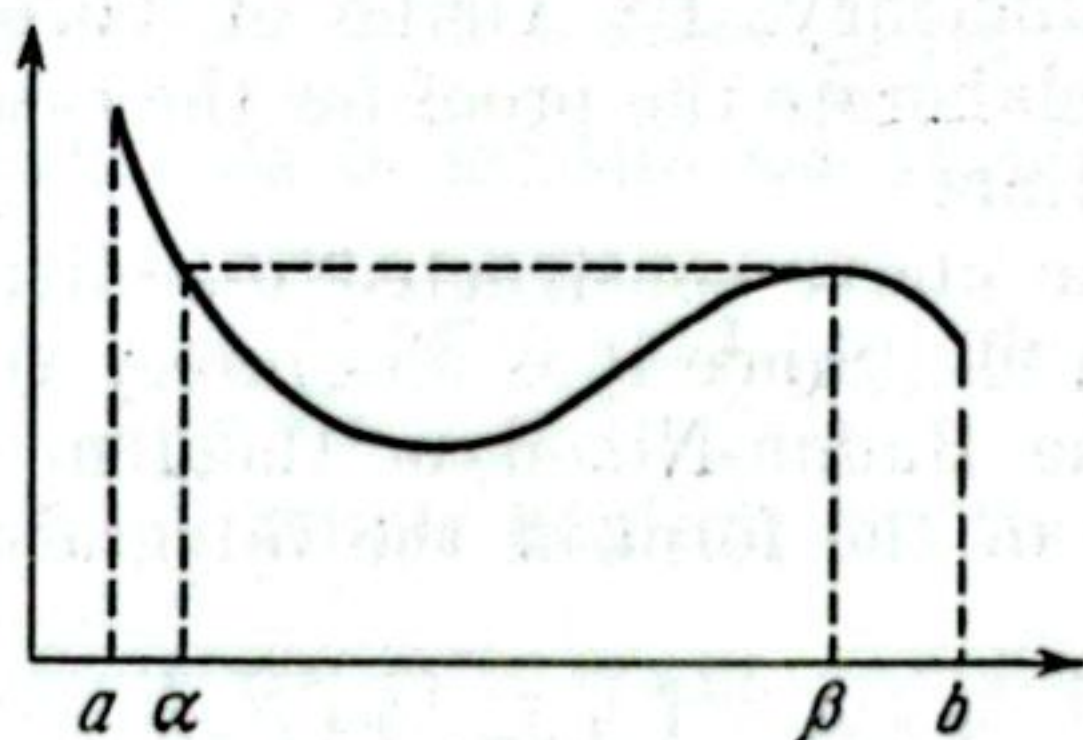


Fig. 21

either in K_r or in K_l . But any other interior point of the interval $[a, b]$ belongs to at least one of these sets. At the same time, one of the sets K_r and K_l can be empty: for example, $K_l = \emptyset$ for every increasing function. For the function whose graph is shown in Fig. 21 we have $K_r = (\alpha, \beta)$ and $K_l = (a, b)$.

Lemma XIII.3.1. *The sets K_r and K_l are open for any continuous function g defined on $[a, b]$. If an open interval (α, β) is a component* of the set K_r (K_l) then $g(\alpha) \leq g(\beta)$ ($g(\alpha) \geq g(\beta)$).*

Proof. We shall consider the set K_r (for K_l the argument is similar). If $x \in K_r$, $\xi > x$ and $g(x) < g(\xi)$ then, by the continuity of the function g , the latter inequality holds in a neighbourhood of the point x and, consequently, x is an interior point of K_r , i.e. K_r is an open set.

Let an open interval (α, β) be a component of the set K_r . Suppose that $g(\alpha) > g(\beta)$. Then there is a point x_0 such

* See Sec. II.4.

that $\alpha < x_0 < \beta$ and $g(x_0) > g(\beta)$. Let x_1 be a point at which the function g attains its greatest value on the interval $[x_0, \beta]$. Obviously, $x_1 \neq \beta$ and therefore $x_1 \in K_r$. Consequently, there is a point $\xi > x_1$ such that $g(\xi) > g(x_1)$, and, by the definition of the point x_1 , the inequality $x_1 < \xi \leq \beta$ is impossible. If we suppose that $\xi > \beta$, the inequalities $g(\beta) < g(x_0) \leq g(x_1) < g(\xi)$ imply that $\beta \in K_r$, which is impossible either. The contradiction we have obtained shows that $g(\alpha) \leq g(\beta)$.*

Now let F be an arbitrary function (with finite values) defined on $[a, b]$ and $x_0 \in (a, b)$. Let us form the ratio

$$\frac{F(x) - F(x_0)}{x - x_0} \quad (5)$$

and state the following definitions: the upper limit as $x \rightarrow x_0$ from the right of this ratio, i.e. $\lim_{x \rightarrow x_0+0} \frac{F(x) - F(x_0)}{x - x_0}$

(the lower limit as $x \rightarrow x_0$ from the right, i.e. $\lim_{x \rightarrow x_0+0} \frac{F(x) - F(x_0)}{x - x_0}$)

will be called the *right-hand upper (lower) derived number of the function F at the point x_0* and will be denoted $\bar{D}_r F(x_0)$ ($\underline{D}_r F(x_0)$). The *left-hand upper and lower derived numbers* are defined similarly and will be denoted, respectively, $\bar{D}_l F(x_0)$ and $\underline{D}_l F(x_0)$. It is clear that the existence of the ordinary derivative $F'(x_0)$ at the point x_0 is equivalent to the equalities

$$\bar{D}_r F(x_0) = \underline{D}_r F(x_0) = \bar{D}_l F(x_0) = \underline{D}_l F(x_0)$$

In this case all the four derived numbers coincide with $F'(x_0)$.

Let us discuss the following example. Consider the function F defined on the interval $[0, 1]$ by means of the relations

$$F(x) = \begin{cases} 1 & \text{if } x \text{ is a rational number} \\ 0 & \text{if } x \text{ is an irrational number} \end{cases}$$

Then, if x_0 is a rational number ratio (5) can only assume negative values or be equal to zero for $x > x_0$, and, as x approaches x_0 from the right, it can become arbitrarily large

* It is easy to understand that if $\alpha = a_1$ then $g(\alpha) = g(\beta)$.

in its absolute value, whence it follows immediately that $\bar{D}_r F(x_0) = 0$ and $\underline{D}_r F(x_0) = -\infty$. On the other hand, for $x < x_0$ ratio (5) is nonnegative, and $\bar{D}_l F(x_0) = +\infty$ while $\underline{D}_l F(x_0) = 0$. If the number x_0 is irrational the situation changes and it can easily be checked that in this case

$$\bar{D}_r F(x_0) = +\infty, \quad \underline{D}_r F(x_0) = \bar{D}_l F(x_0) = 0, \\ \underline{D}_l F(x_0) = -\infty$$

The function under consideration has neither an ordinary (two-sided) derivative nor even one-sided derivatives (i.e. the left-hand and the right-hand derivatives) at every point of the domain of definition of the function.

If F is an increasing function ratio (5) is nonnegative for any $x \neq x_0$ and therefore all the four derived numbers of F are also nonnegative. In what follows we shall suppose that the function F is *continuous* and *increasing* throughout the interval $[a, b]$.

Lemma XIII.3.2. *Let $E \subset (a, b)$ be an arbitrary set and let there exist a constant $M > 0$ such that $\bar{D}_r F(x) > M$ at every point $x \in E$. Then*

$$\mu^* E \leq \frac{1}{M} [F(b) - F(a)] \quad (6)$$

Proof. It follows from the definition of $\bar{D}_r F$ that for any $x_0 \in E$ there is $\xi > x_0$ such that

$$\frac{F(\xi) - F(x_0)}{\xi - x_0} > M$$

whence

$$F(\xi) - M\xi > F(x_0) - Mx_0 \quad (7)$$

Let us take the function

$$g(x) = F(x) - Mx$$

which is continuous throughout the interval $[a, b]$. Inequality (7) means that x_0 is an r -point of the function g . Thus, $E \subset K_r(g)$.

The foregoing lemma indicates that $K_r(g)$ is an open set. Consequently, it is representable as the union of its components:

$$K_r(g) = \bigcup_n (\alpha_n, \beta_n)$$

Furthermore, we have $g(\alpha_n) \leq g(\beta_n)$ for each n , that is

$$M(\beta_n - \alpha_n) \leq F(\beta_n) - F(\alpha_n)$$

whence

$$\begin{aligned} \mu^* E &\leq \mu K_r(g) = \sum_n (\beta_n - \alpha_n) \leq \\ &\leq \frac{1}{M} \sum_n [F(\beta_n) - F(\alpha_n)] \leq \frac{1}{M} [F(b) - F(a)] \end{aligned}$$

Corollary. If $E = \{x; x \in (a, b), \overline{D}_r F(x) = +\infty\}$ then $\mu E = 0$.

Indeed, in this case inequality (6) holds for any $M > 0$.

Lemma XIII.3.3. Let $E \subset (a, b)$ be a set for which there is a constant $N > 0$ such that $\underline{D}_l F(x) < N$ at each point $x \in E$. Then there exists an open set $G \subset (a, b)$ such that $E \subset G$ and the inequality

$$F(\beta) - F(\alpha) \leq N(\beta - \alpha) \quad (8)$$

holds for every component (α, β) of the set G .

Proof. Let $x_0 \in E$. By the definition of $\underline{D}_l F$, there is $\xi < x_0$ such that

$$\frac{F(\xi) - F(x_0)}{\xi - x_0} < N$$

whence

$$F(\xi) - N\xi > F(x_0) - Nx_0$$

By analogy with the foregoing proof, we put $g(x) = F(x) - Nx$ and conclude that $x_0 \in K_l(g)$. Hence, if $G = K_l(g)$ we have $E \subset G$, and G is an open set. Moreover, for any component (α, β) of G there must be $g(\alpha) \geq g(\beta)$, which is equivalent to inequality (8).

Lemma XIII.3.4. Let $0 < N < M < +\infty$, $E \subset (a, b)$ and

$$\underline{D}_l F(x) < N < M < \overline{D}_r F(x)$$

for each $x \in E$. Then $\mu E = 0$.

Proof. Let G be an open set satisfying all the requirements of the foregoing lemma and let (α_n, β_n) ($n = 1, 2, \dots$) be its components. If we put $E_n = E \cap (\alpha_n, \beta_n)$ then $E = \bigcup_n E_n$. Considering the function F on the interval $[\alpha_n,$

$\beta_n]$ we can apply Lemma XIII.3.2 to each E_n ($n = 1, 2, \dots$). By this lemma and inequality (8),

$$\mu^* E_n \leq \frac{1}{M} [F(\beta_n) - F(\alpha_n)] \leq \frac{N}{M} (\beta_n - \alpha_n)$$

Whence, putting $\lambda = \frac{N}{M}$ ($0 < \lambda < 1$), we obtain

$$\mu^* E \leq \sum_n \mu^* E_n \leq \lambda \sum_n (\beta_n - \alpha_n) = \lambda \mu G \leq \lambda (b - a)$$

Let us take an arbitrary interval $[a', b'] \subset [a, b]$ and put $E' = E \cap (a', b')$. Applying what has already been proved to the set E' we obtain $\mu^* E' \leq \lambda (b' - a')$. Now it only remains to derive from this result the equality $\mu E = 0$.

By Theorem V.5.5, given any $\varepsilon > 0$, there is an open set $G \supset E$ such that

$$\mu G < \mu^* E + \varepsilon$$

Without losing generality we can assume that $G \subset (a, b)$. Let $G = \bigcup_n (a_n, b_n)$ be the representation of the set G as the union of its components. Then $E = \bigcup_n E_n$ where $E_n = E \cap (a_n, b_n)$ and, by the above,

$$\mu^* E \leq \sum_n \mu^* E_n \leq \lambda \sum_n (b_n - a_n) = \lambda \mu G < \lambda (\mu^* E + \varepsilon)$$

On passing to the limit for $\varepsilon \rightarrow 0$ we obtain $\mu^* E \leq \lambda \mu^* E$, which is only possible if $\mu^* E = 0$.

Lemma XIII.3.5. *Let*

$$E = \{x; x \in (a, b), \underline{D}_l F(x) < \overline{D}_r F(x)\}$$

Then $\mu E = 0$.

Proof. Let us consider all the possible pairs of positive rational numbers p and q such that $p < q$. These pairs form a countable set. Let

$$E_{p,q} = \{x; x \in (a, b), \underline{D}_l F(x) < p < q < \overline{D}_r F(x)\}$$

It is clear that every $E_{p,q} \subset E$ and, conversely, for any $x \in E$ there is a pair of rational numbers p and q such that $x \in E_{p,q}$. Thus, $E = \bigcup_{p < q} E_{p,q}$. According to the preceding lemma, we have $\mu E_{p,q} = 0$ for any pair p, q , and thus $\mu E = 0$.

Theorem XIII.3.1. (H. Lebesgue). *If F is a continuous increasing function on $[a, b]$ then the derivative $F'(x)$ exists and is finite almost everywhere on $[a, b]$.*

Proof. The corollary of Lemma XIII.3.2 implies that $\bar{D}_r F(x) < +\infty$ almost everywhere and Lemma XIII.3.5 indicates that $\underline{D}_l F(x) \geq \bar{D}_r F(x)$ also almost everywhere. Let us consider the function $\Phi(x) = -F(-x)$ defined on

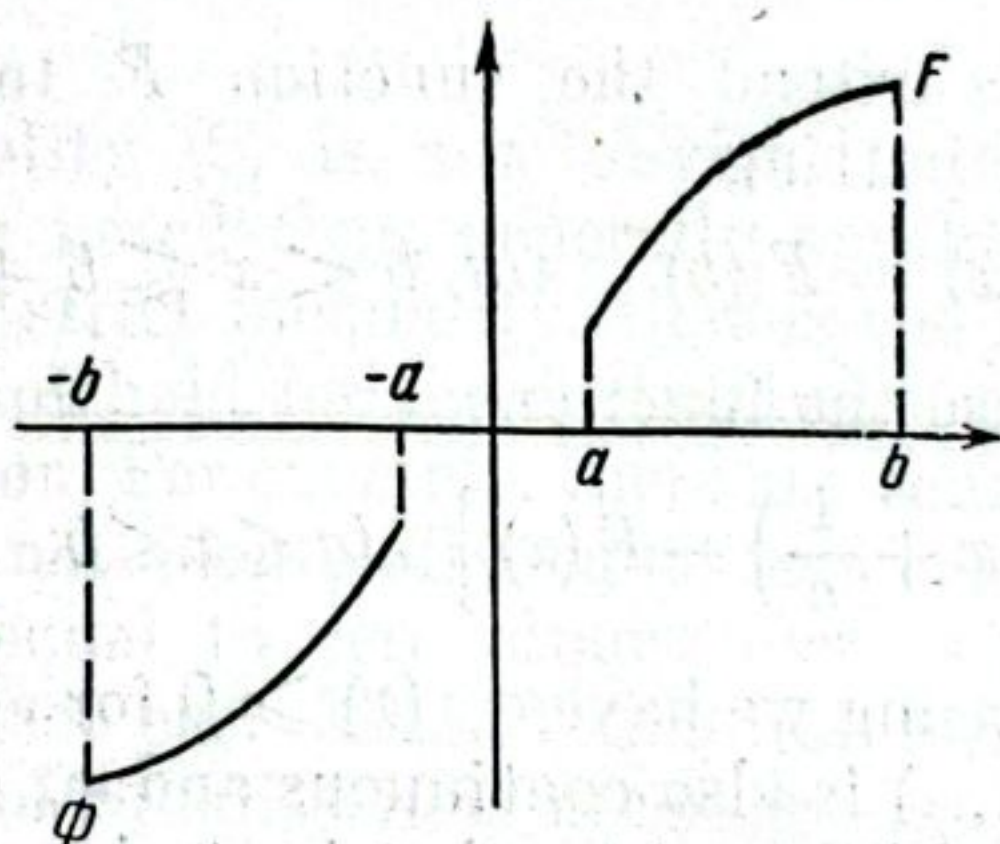


Fig. 22

the interval $[-b, -a]$. Like F , this function is continuous and increasing (see Fig. 22). Consequently, for $\Phi(x)$ we also have

$$\underline{D}_l \Phi(x) \geq \bar{D}_r \Phi(x) \text{ almost everywhere}$$

At the same time, it follows from the equality

$$\frac{F(x) - F(x_0)}{x - x_0} = \frac{\Phi(-x) - \Phi(-x_0)}{-x - (-x_0)}$$

that $\bar{D}_l F(x_0) = \bar{D}_r \Phi(-x_0)$ and $\underline{D}_r F(x_0) = \underline{D}_l \Phi(-x_0)$, and therefore $\underline{D}_r F(x) \geq \bar{D}_l F(x)$ almost everywhere. Furthermore, by definition,

$$\underline{D}_r F(x) \leq \bar{D}_r F(x) \text{ and } \underline{D}_l F(x) \leq \bar{D}_l F(x)$$

for all x . Combining the inequalities obtained for the derived numbers of the function F we see that

$$0 \leq \bar{D}_r F(x) \leq \underline{D}_l F(x) \leq \bar{D}_l F(x) \leq \underline{D}_r F(x) \leq \bar{D}_r F(x) < +\infty$$

almost everywhere on the interval $[a, b]$ and, consequently, the derivative $F'(x)$ exists and is finite almost everywhere.

Theorem XIII.3.2. *If F is a continuous increasing function on an interval $[a, b]$ its derivative F' is summable on that interval and*

$$\int_a^b F' dx \leq F(b) - F(a) \quad (9)$$

Proof. Let us extend the function F to the interval $[a, b+1]$ by putting

$$F(x) = F(b) \quad \text{for } b < x \leq b+1$$

We shall also use the functions

$$\varphi_n(x) = n \left[F\left(x + \frac{1}{n}\right) - F(x) \right] \quad (a \leq x \leq b; \quad n = 1, 2, \dots)$$

Since F is increasing we have $\varphi_n(x) \geq 0$ for all x . Moreover, φ_n ($n = 1, 2, \dots$) is also continuous and $\varphi_n(x) \rightarrow F'(x)$ at all points where $F'(x)$ exists, that is at almost all points of the interval $[a, b]$. By Theorem VI.4.2, it follows that F' is measurable. Using Fatou's theorem we can write

$$\int_a^b F' dx \leq \sup_n \int_a^b \varphi_n dx \quad (10)$$

But

$$\begin{aligned} \int_a^b \varphi_n dx &= n \left[\int_a^b F\left(x + \frac{1}{n}\right) dx - \int_a^b F(x) dx \right] = \\ &= n \left[\int_{a+\frac{1}{n}}^{b+\frac{1}{n}} F(x) dx - \int_a^b F(x) dx \right]^* = \\ &= n \left[\int_b^{b+\frac{1}{n}} F(x) dx - \int_a^{a+\frac{1}{n}} F(x) dx \right] = F(b) - n \int_a^{a+\frac{1}{n}} F(x) dx \end{aligned}$$

* Here we have used the classical rule for change of variable in the definite integral of a continuous function.

and, since $F(x) \geq F(a)$, we have

$$n \int_a^{a + \frac{1}{n}} F(x) dx \geq F(a)$$

and hence

$$\int_a^b \varphi_n dx \leq F(b) - F(a) \quad (n = 1, 2, \dots)$$

Lastly, inequality (9) is now obviously implied by (10).

It should be noted that, generally speaking, formula (9) may involve a strict inequality, that is the Newton-Leibniz formula may not hold for the derivative of a continuous increasing function. For example, there are continuous increasing functions not identically equal to a constant whose derivatives are equal to zero almost everywhere*. Later on we shall show (see Sec. XIII.5) that for formula (9) to involve the sign of equality it is necessary and sufficient that the function F should be absolutely continuous.

To conclude this section we shall present a theorem on termwise differentiation of a series of monotone functions.

Theorem XIII.3.3 (G. Fubini). *If a series $\sum_{n=1}^{\infty} F_n$ whose members $F_n(x)$ are continuous increasing functions defined on an interval $[a, b]$ is convergent at all the points of that interval to a function F (i.e. $F(x) = \sum_{n=1}^{\infty} F_n(x)$) then, for almost all $x \in [a, b]$, it is allowable to differentiate the series termwise, that is there exists a finite derivative $F'(x)$ and*

$$F'(x) = \sum_{n=1}^{\infty} F'_n(x) \quad (11)$$

Proof. Here we shall present a proof of the theorem based on Theorem XIII.3.1 and therefore it should be additionally assumed that all the functions F_n and F are continuous. In the next section we shall show that the theorem continues to hold without this assumption.

* E.g. see I. P. Natanson, *The Theory of Functions of a Real Variable*, Fizmatgiz, Moscow, 1957 (see the bibliography).

(a) First of all, we readily see that F is also an increasing function. Furthermore, without loss of generality, we can assume that $F_n(a) = 0$ for all n . Let us check that the series in formula (11) is convergent almost everywhere. It follows from Theorem XIII.3.1 that all the functions F_n and the function F possess, respectively, finite derivatives $F'_n(x)$ and $F'(x)$ for almost all $x \in [a, b]$. Let us denote the set of all such x 's by A . Putting

$$S_n(x) = \sum_{k=1}^n F_k(x) \quad (n = 1, 2, \dots)$$

we see that $S_{n+1} - S_n$ and also $F - S_n$ are increasing functions whence it follows that for all $x \in A$ we have

$$S'_n(x) \leq S'_{n+1}(x) \leq F'(x)$$

Consequently, there exists a finite limit of the functional sequence $\{S'_n(x)\}$, that is the series $\sum_{n=1}^{\infty} F'_n(x)$ is convergent for all $x \in A$.

(b) Now we proceed to prove equality (11). Let us choose an increasing sequence of indices $\{n_k\}$ such that

$$F(b) - S_{n_k}(b) < \frac{1}{k^2} \quad (k = 1, 2, \dots)$$

Then

$$(i) \quad 0 \leq F(x) - S_{n_k}(x) \leq F(b) - S_{n_k}(b) < \frac{1}{k^2} \text{ for any } x \in [a, b];$$

$$(ii) \quad \text{the series } \sum_{k=1}^{\infty} [F(b) - S_{n_k}(b)] \text{ is convergent.}$$

Therefore the series

$$\sum_{k=1}^{\infty} [F(x) - S_{n_k}(x)]$$

is uniformly convergent on the interval $[a, b]$ and, since its terms are continuous, its sum is also continuous. The terms of the last series being continuous increasing functions, it satisfies the conditions of the theorem and, according to what has been proved in (a), the series

$$\sum_{k=1}^{\infty} [F'(x) - S'_{n_k}(x)]$$

composed of the derivatives converges almost everywhere. Consequently,

$$F'(x) - S'_{n_k}(x) \xrightarrow[k \rightarrow \infty]{} 0 \text{ almost everywhere}$$

that is

$$S'_{n_k}(x) \xrightarrow[k \rightarrow \infty]{} F'(x) \text{ almost everywhere}$$

But since the sequence $\{S'_n(x)\}$ is increasing for every $x \in A$ its limit (finite or equal to $+\infty$) exists and must coincide with the limiting of its subsequence. Hence, $S'_n(x) \rightarrow F'(x)$ almost everywhere, which implies equality (11).

It should be stressed that the continuity of the functions F_n and F was only used in this proof to guarantee the existence almost everywhere of finite derivatives $F'_n(x)$ and $F'(x)$.

§ 4. Differentiation of Discontinuous Monotone Functions

Here we shall extend the results of the foregoing section to arbitrary monotone functions. Let F be an increasing function defined on an interval $[a, b]$. As is known, for such a function there exist at each point $x_0 \in (a, b)$ the finite right-hand and left-hand limits $F(x_0 + 0)$ and $F(x_0 - 0)^*$ and

$$F(x_0 - 0) \leq F(x_0) \leq F(x_0 + 0)$$

The points of discontinuity of the function F are those for which at least one of these inequalities is strict. If the end point a (b) is a point of discontinuity, then $F(a) < F(a + 0)$ ($F(b - 0) < F(b)$). The difference

$$\delta_+(x_0) = F(x_0 + 0) - F(x_0)$$

where x_0 is an arbitrary point of the interval (a, b) is called the *right-hand jump* of the function F at the point x_0 , the difference

$$\delta_-(x_0) = F(x_0) - F(x_0 - 0)$$

is called the *left-hand jump* of F at x_0 and the difference

$$\delta(x_0) = F(x_0 + 0) - F(x_0 - 0)$$

* A monotone function can only have discontinuities of the first kind.

is simply called the *jump* of F at x_0 . At the point a only the right-hand jump makes sense, and we put, by definition, $\delta(a) = \delta_+(a)$. Similarly, at the point b there exists only the left-hand jump, and we put $\delta(b) = \delta_-(b)$. The continuity of the function F from the left or from the right at a given point means that the corresponding jump is equal to zero.

Theorem XIII.4.1. *The set H of the points of discontinuity of an increasing function F is at most countable.*

Proof. Let F be defined on an interval $[a, b]$, H be the set of all the points of discontinuity of F , and H_n be the set of all the points at which $\delta(x) > \frac{1}{n}$ ($n = 1, 2, \dots$). Then

$H = \bigcup_{n=1}^{\infty} H_n$. We shall show that each of the sets H_n is finite, which will imply that H is at most countable.

Let us put $A = F(b) - F(a)$ and show that the number of points contained in H_n must be less than nA . Suppose that H_n contains k distinct points. Let them be numbered in an increasing order:

$$a \leq c_1 < c_2 < \dots < c_k \leq b$$

Then

$$F(a) \leq F(c_1 - 0) < F(c_1 + 0) \leq F(c_2 - 0) < F(c_2 + 0) \leq \dots \leq F(c_k - 0) < F(c_k + 0) \leq F(b)^*$$

whence

$$\sum_{i=1}^k [F(c_i + 0) - F(c_i - 0)] \leq F(b) - F(a) = A \quad (12)$$

Since $c_i \in H_n$ we have

$$F(c_i + 0) - F(c_i - 0) = \delta(c_i) > \frac{1}{n}$$

and, consequently, $k \cdot \frac{1}{n} < A$, that is $k < nA$. The theorem has been proved.

Now let us renumber the points of the set H , that is the points of discontinuity of the function F , arranging them

* If $c_1 = a$ ($c_k = b$) then $F(c_1 - 0)$ ($F(c_k + 0)$) should be omitted. Accordingly, in the sum in inequality (12) it should be replaced by $F(a)$ ($F(b)$).

into a single sequence $H = \{c_j\}$. Then

$$\sum_j \delta(c_j) \leq A = F(b) - F(a) \quad (13)$$

Indeed, for any finite number of summands on the left-hand side inequality (13) is obtained by analogy with (12), and if the set H is countable we arrive at the same estimation by means of the passage to the limit.

Proceeding from the function F , let us define a function S throughout the interval $[a, b]$ by means of the relations

$$S(a) = 0$$

and

$$S(x) = \sum_{j, c_j < x} \delta(c_j) + \delta_-(x) \text{ for } a < x \leq b$$

(here the summation extends over all indices j for which the points of discontinuity c_j lie strictly on the left of x). The function S is termed the *jump function* of the function F . It is clear that S is an increasing function and $0 \leq S(x) \leq S(b) \leq A$ for all x .

Theorem XIII.4.2. *The difference $G = F - S$ is a continuous increasing function on the interval $[a, b]$.*

Proof. Let $a \leq x_1 < x_2 \leq b$. The difference $S(x_2) - S(x_1)$ is equal to $\delta_+(x_1) + \sum_{j(x_1 < c_j < x_2)} \delta(c_j) + \delta_-(x_2)$

and therefore it is nothing but the sum of all the jumps of the restriction of the function F to the interval $[x_1, x_2]$. Applying formula (13) to this interval we find that

$$S(x_2) - S(x_1) \leq F(x_2) - F(x_1) \quad (14)$$

whence $G(x_2) - G(x_1) \geq 0$. Consequently, G is an increasing function.

Let us take an arbitrary point $x_0 \in [a, b)$ and show that the function G is continuous from the right at this point. If $x > x_0$ formula (14) shows that

$$S(x) - S(x_0) \leq F(x) - F(x_0)$$

On passing to the limit in this relation as $x \rightarrow x_0 + 0$ we obtain

$$S(x_0 + 0) - S(x_0) \leq F(x_0 + 0) - F(x_0) \quad (15)$$

On the other hand,

$$S(x) - S(x_0) \geq \delta_+(x_0) = F(x_0 + 0) - F(x_0)$$

and therefore we also have

$$S(x_0 + 0) - S(x_0) \geq F(x_0 + 0) - F(x_0) \quad (16)$$

Comparing (15) and (16) we see that

$$S(x_0 + 0) - S(x_0) = F(x_0 + 0) - F(x_0)$$

whence

$$G(x_0 + 0) = F(x_0 + 0) - S(x_0 + 0) = F(x_0) - S(x_0) = G(x_0)$$

The continuity of the function G from the left is proved in like manner.

Now we proceed to generalize Theorem XIII.3.1 to arbitrary increasing functions.

Let us consider the class of functions g defined on an interval $[a, b]$ whose all discontinuities (provided there are such) are of the first kind (jump discontinuities). In particular, this class contains all functions continuous on the interval $[a, b]$ and all monotone functions. Note that the sum of two functions belonging to this class also belongs to it.

If a function g has only jump discontinuities we put

$$G(x) = \max [g(x), g(x + 0), g(x - 0)]$$

for every point $x_0 \in (a, b)$ and

$$G(a) = \max [g(a), g(a + 0)],$$

$$G(b) = \max [g(b), g(b - 0)]$$

for the end points of the interval $[a, b]$.

It can be proved quite simply that every function belonging to the class under consideration is bounded and that its supremum "is attained" in the following sense: if $M = \sup g(x)$ on $[a, b]$ then there exists a point $x_0 \in [a, b]$ at which $G(x_0) = M$.

The definition of r -points and l -points stated in the foregoing section for continuous functions is extended to functions with discontinuities of the first kind as follows: $x_0 \in (a, b)$ is an r -point (l -point) if there is $\xi > x_0$ ($\xi < x_0$) such that $G(x_0) < g(\xi)$.

The generalization of Lemma XIII.3.1 to the class of functions with jump discontinuities reads as follows: if g is a function belonging to this class on an interval $[a, b]$, the set K_r of its r -points and the set K_l of its l -points are open; if an open interval (α, β) is a component of the set K_r (K_l) then $g(\alpha + 0) \leq G(\beta)$ ($G(\alpha) \geq g(\beta - 0)$).

To prove that K_r is an open set it is sufficient to note that if $G(x_0) < g(\xi)$ the inequality $g(x) < c < g(\xi)$ is fulfilled for all x belonging to a neighbourhood of the point x_0 , and then $G(x) < g(\xi)$ in that neighbourhood. The considerations concluding the proof are almost the same as in Lemma XIII.3.1, and we leave them to the reader.

Now, let F be an arbitrary increasing function on an interval $[a, b]$. Lemma XIII.3.2 continues to hold for it. Since the set H of all the points of discontinuity of the function F is at most countable and thus is of measure zero it is sufficient to derive estimation (6) for the set $E \setminus H$. For the latter set the proof of the lemma remains almost the same*. Furthermore, noting that the values of the function F at the end points of the interval $[a, b]$ do not affect its derived numbers at the interior points we can replace inequality (6) by the more precise inequality

$$\mu^* E \leq \frac{1}{M} [F(b - 0) - F(a + 0)] **$$

In Lemma XIII.3.3 we obtain, instead of inequality (8), the relation

$$F(\beta - 0) - F(\alpha + 0) \leq N(\beta - \alpha) \quad (17)$$

On the set G we should impose the requirement that $E \setminus H \subset G$.

After this we can prove without any changes Lemmas XIII.3.4 and XIII.3.5 and then Theorem XIII.3.1 as well. Thus, every increasing function F defined on an interval $[a, b]$ possesses a finite derivative $F'(x)$ almost everywhere on that

* In the course of the proof it should be taken into account that, since F is an increasing function and $g(x) = F(x) - Mx$ (all discontinuities of g are of the first kind), we have $G(x) = F(x + 0) - Mx$ for any $x < b$.

** To derive this relation we temporarily regard $F(a + 0)$ and $F(b - 0)$ as the values of the function F at the end points of the interval $[a, b]$.

interval. The same obviously applies to any function of bounded variation.

After Theorem XIII.3.1 has been extended to arbitrary increasing functions, Fubini's theorem (XIII.3.3) can also be generalized to series composed of arbitrary increasing functions. This facilitates the investigation of differentiability property of a jump function.

Theorem XIII.4.3. *Let F be an increasing function on an interval $[a, b]$ and S be its jump function. Then $S'(x) = 0$ almost everywhere on $[a, b]$.*

Proof. Let us number all the points of discontinuity of the function F and arrange them into a sequence $c_1, c_2, \dots$. To each point $c_k \in (a, b)$ we assign the function defined on the interval $[a, b]$ by means of the formulas

$$S_k(x) = \begin{cases} 0 & \text{for } x < c_k \\ \delta_-(c_k) & \text{for } x = c_k \\ \delta(c_k) & \text{for } x > c_k \end{cases}$$

If $c_k = a$ or $c_k = b$ for a certain k we put, respectively,

$$S_k(x) = \begin{cases} 0 & \text{for } x = a \\ \delta(a) & \text{for } x > a \end{cases}$$

or

$$S_k(x) = \begin{cases} 0 & \text{for } x < b \\ \delta(b) & \text{for } x = b \end{cases}$$

It is clear that each of the functions S_k ($k = 1, 2$) is increasing and that $S'_k(x) = 0$ for almost all $x \neq c_k$. Furthermore,

$$S(x) = \sum_k S_k(x)$$

for all $x \in [a, b]$. Therefore Fubini's theorem* implies that

$$S'(x) = \sum_k S'_k(x) = 0 \text{ almost everywhere}$$

Now we shall show that Theorem XIII.3.2 continues to hold for an arbitrary increasing function F .** Let S be the

* It is clear that the reference to Fubini's theorem is only necessary if the function F has an infinite number of discontinuities.

** By the way, the proof of Theorem XIII.3.2 in the foregoing section should only be slightly modified to apply to the general case.

jump function of F and $G = F - S$. Then, by Theorem XIII.4.2, G is a continuous increasing function and therefore, as was proved in Theorem XIII.3.2,

$$\int_a^b G' dx \leq G(b) - G(a)$$

Furthermore, by the foregoing theorem, there must be $F'(x) = G'(x)$ almost everywhere on $[a, b]$ and

$$\begin{aligned} F(b) - F(a) &= G(b) - G(a) + S(b) - S(a) \geq \\ &\geq G(b) - G(a) \end{aligned}$$

Consequently,

$$\int_a^b F'(x) dx \leq F(b) - F(a)$$

§ 5. Differentiation of the Integral

In this section we shall answer the second of the questions posed at the beginning of the present chapter, that is we shall find out whether there holds the equality $F'(x) = f(x)$ if the function F is representable in the form of an integral of a summable function f :

$$F(x) = \int_a^x f dx$$

As in the classical case, it can be proved that $F'(x) = f(x)$ at each point of continuity of the function f . But we know that a summable function may have no points of continuity. We shall see that it can nevertheless be shown that there are "very many" points at which the equality $F'(x) = f(x)$ holds for an arbitrary summable function f .

Lemma XIII.5.1. *If f is a summable function on an interval $[a, b]$ and the inequality*

$$\int_{\alpha}^{\beta} f dx \geq 0$$

holds for any interval $[\alpha, \beta] \subset [a, b]$ then $f(x) \geq 0$ almost everywhere on $[a, b]$.

Proof. The conditions of the lemma readily show that

$$\int_G f_1 d\mu \geq 0,$$

for any open set $G \subset (a, b)$. If $K \subset (a, b)$ is a G_δ -set it is representable as the intersection of a decreasing sequence of open sets $G_n \subset (a, b)$: $K = \bigcap_{n=1}^{\infty} G_n$. Then, by Theorem IV.1.2,

$$\int_K f d\mu = \lim \int_{G_n} f d\mu \geq 0$$

Lastly, let E be an arbitrary measurable subset of (a, b) . By Theorem V.5.6, there exists a G_δ -set $K \subset (a, b)$ such that $E \subset K$ and $\mu(K \setminus E) = 0$. Therefore

$$\int_E f_1 d\mu = \int_K f d\mu \geq 0$$

If we suppose that $f(x) < 0$ on a set E of positive measure (it is allowable to assume that $E \subset (a, b)$) then, by Theorem XIII.3.6, $\int_E f d\mu < 0$, we thus arrive at a contradiction.

Theorem XIII.5.1. *Let f be a summable function on $[a, b]$ and*

$$F(x) = \int_a^x f dx \quad (18)$$

Then $F'(x) = f(x)$ almost everywhere on $[a, b]$.

Proof. Since every summable function can be represented as a difference of two nonnegative summable functions it suffices to prove the theorem for the case when $f(x) \geq 0$.

We shall begin with a bounded function f : $0 \leq f(x) \leq M$.

Let us put $g(x) = M - f(x)$ and $G(x) = \int_a^x g dx$. Then

$$H(x) \equiv \int_a^x (f + g) dx = M(x - a)$$

and the function H possesses the derivative $H'(x) = M$ for all x .

Both functions F and G are increasing (and continuous) and, by Theorem XIII.3.2,

$$\int_{\alpha}^{\beta} F' dx \leq F(\beta) - F(\alpha) = \int_{\alpha}^{\beta} f dx$$

for any interval $[\alpha, \beta] \subset [a, b]$. Consequently,

$$\int_{\alpha}^{\beta} (f - F') dx \geq 0$$

Therefore, by the preceding lemma, $f(x) - F'(x) \geq 0$ almost everywhere on $[a, b]$. Completely similarly, $g(x) - G'(x) \geq 0$ almost everywhere. If at least one of these relations involves a strict inequality on a set of positive measure there is at least one point x such that

$$M = f(x) + g(x) > F'(x) + G'(x) = H'(x) = M^*$$

and we thus arrive at a contradiction. Consequently, $F'(x) = f(x)$ almost everywhere.

Now we pass to the case when f is an unbounded summable function (as before, we assume that $f(x) \geq 0$). Let f_m ($m = 1, 2, \dots$) be the corresponding truncated functions and $F_m(x) = \int_a^x f_m dx$. By what has already been proved, $F'_m = f_m(x)$ almost everywhere. Let us form the series

$$F_1 + (F_2 - F_1) + \dots + (F_m - F_{m-1}) + \dots$$

Since $f_m(x) - f_{m-1}(x) \geq 0$ for all x the functions $F_m - F_{m-1}$ are increasing. The sum of this series is equal to

$$\lim F_m(x) = \lim \int_a^x f_m dx = \int_a^x f dx = F(x)$$

* If, for example, $f(x) > F'(x)$ on a set of positive measure this set must necessarily have a nonempty intersection with the set on which $f(x) \geq G'(x)$ since the latter inequality holds almost everywhere.

(see proposition (a) in Sec. VIII.4) and, by Theorem XIII.3.3*,

$$\begin{aligned} F'(x) &= F'_1(x) + \sum_{m=2}^{\infty} [F'_m(x) - F'_{m-1}(x)] = \\ &= f_1(x) + \sum_{m=2}^{\infty} [f_m(x) - f_{m-1}(x)] = f(x) \end{aligned}$$

almost everywhere on $[a, b]$. The theorem has been proved.

Combining the result established here with Theorem XIII.2.2 we conclude that *every absolutely continuous function F can be represented by means of formula (4) as the integral of its derivative!*

$$F(x) = F(a) + \int_a^x F' dx$$

This immediately implies the following propositions:

1°. If F is absolutely continuous and $F'(x) = 0$ almost everywhere then $F(x) = \text{const.}$

2°. If F and G are absolutely continuous and $F'(x) = G'(x)$ almost everywhere then $F(x) - G(x) = \text{const.}$

Let us return to an arbitrary increasing function F and to formula (9).

Theorem XIII.5.2. *For formula (9) to involve the sign of equality it is necessary and sufficient that the function F be absolutely continuous.*

Proof. The sufficiency of this condition follows from the foregoing theorem. Let us check its necessity.

Let

$$\int_a^b F' dx = F(b) - F(a) \quad (19)$$

Applying formula (9) to the intervals $[a, x]$ and $[x, b]$ where $x \in (a, b)$ is an arbitrary point we obtain

$$\int_a^x F' dx \leq F(x) - F(a) \quad \text{and} \quad \int_x^b F' dx \leq F(b) - F(x)$$

* Note that here we apply Fubini's theorem to a series composed of continuous functions and therefore the reference to the generalization of this theorem given in Sec. XIII.4 is not needed.

If at least one of these formulas involved the sign of strict inequality we should obtain

$$\int_a^b F' dx < F(b) - F(a)$$

which contradicts condition (19). Consequently, the function F is representable by formula (4) and therefore it is absolutely continuous.

We shall also present a strengthened variant of Theorem XIII.5.1. The relation $F'(x_0) = f(x_0)$ means, for function (18), that

$$\frac{1}{h} \int_{x_0}^{x_0+h} f dx \xrightarrow{h \rightarrow 0} f(x_0)$$

that is (if $f(x_0)$ is finite),

$$\frac{1}{h} \int_{x_0}^{x_0+h} [f(x) - f(x_0)] dx \xrightarrow{h \rightarrow 0} 0 \quad (20)$$

Strengthening relation (20) we state the following definition:

Definition. A point $x_0 \in [a, b]$ is called a *Lebesgue point* of a function f summable on the interval $[a, b]$ if $f(x_0)$ is finite and

$$\frac{1}{h} \int_{x_0}^{x_0+h} |f(x) - f(x_0)| dx \xrightarrow{h \rightarrow 0} 0 \quad (21)$$

It is readily seen that every point x_0 at which the function f is continuous is its Lebesgue point.

Theorem XIII.5.3. If f is a summable function on an interval $[a, b]$ almost all the points of that interval are the Lebesgue points of the function f .

Proof. Let r be an arbitrary rational number. Since the function f is summable on $[a, b]$, so is the function $|f(x) - r|$. Consequently, by Theorem XIII.5.1,

$$\frac{1}{h} \int_{x_0}^{x_0+h} |f(x) - r| dx \xrightarrow{h \rightarrow 0} |f(x_0) - r|$$

for almost all $x_0 \in [a, b]$. Let $E(r)$ denote the set of all points $x_0 \in [a, b]$ for which the last relation is violated. Its measure $\mu E(r)$ is equal to zero.

Numbering all rational numbers in a certain way $(r_1, r_2, \dots, r_n, \dots)$ we put

$$E = \bigcup_{n=1}^{\infty} E(r_n) \cup \{x; x \in [a, b], f(x) = \pm \infty\}$$

Then $\mu E = 0$. We shall show that every point $x_0 \notin E$ is a Lebesgue point of the function f .

Let $x_0 \notin E$. Given an arbitrary $\varepsilon > 0$, there is a rational number r_n such that

$$|f(x_0) - r_n| < \frac{\varepsilon}{3} \quad (22)$$

Then for any point x at which $f(x)$ is finite we have

$$||f(x) - r_n| - |f(x) - f(x_0)|| < \frac{\varepsilon}{3}$$

The integration yields

$$\left| \frac{1}{h} \int_{x_0}^{x_0+h} |f(x) - r_n| dx - \frac{1}{h} \int_{x_0}^{x_0+h} |f(x) - f(x_0)| dx \right| \leq \frac{\varepsilon}{3}$$

But $x_0 \notin E(r_n)$ and therefore

$$\left| \frac{1}{h} \int_{x_0}^{x_0+h} |f(x) - r_n| dx - |f(x_0) - r_n| \right| < \frac{\varepsilon}{3}$$

for $|h| < \delta(\varepsilon)$. Consequently, by virtue of (22),

$$\left| \frac{1}{h} \int_{x_0}^{x_0+h} |f(x) - r_n| dx \right| < \frac{2}{3} \varepsilon$$

whence

$$\left| \frac{1}{h} \int_{x_0}^{x_0+h} |f(x) - f(x_0)| dx \right| < \varepsilon$$

for $|h| < \delta(\varepsilon)$. We have thus proved relation (21).

Theorem XIII.5.1 on the differentiation of the integral of a summable function can be in a certain sense generalized to integrals in many-dimensional Euclidean spaces. For instance, see I. P. Natanson, *op. cit.*

§ 6. Criterion of F. Riesz

There are some more precise variants of Theorem XIII.2.2 characterizing not general summable functions but some of their special classes. In this section we shall present such a modification of Theorem XIII.2.2 for the class L^p ($p > 1$). To this end we begin with the following

Definition. Let $p > 1$. We shall say that a function F of one variable x defined on an interval $[a, b]$ is of class A^p (i.e. $F \in A^p$) if there exists a constant $K \geq 0$ such that the inequality

$$\sum_{i=0}^{n-1} \frac{|F(x_{i+1}) - F(x_i)|^p}{(x_{i+1} - x_i)^{p-1}} \leq K \quad (23)$$

holds for any partition τ of the interval $[a, b]$ into a finite number of subintervals with the aid of a system of points of division $a = x_0 < x_1 < \dots < x_n = b$.

If $p = 1$ this definition turns into the definition of functions of bounded variation.

In the considerations below we shall make use of Hölder's inequality for sums which is an analogue of Hölder's inequality for integrals established in Sec. X.1. This modification of Hölder's inequality can be derived independently of its integral form, by applying an analogous technique, but here we shall deduce it as a direct consequence of Hölder's inequality in integral form. Let c_i and d_i ($i = 1, 2, \dots, n$) be arbitrary real numbers; we shall consider $p > 1$ and the corresponding conjugate exponent q ($\frac{1}{p} + \frac{1}{q} = 1$). On the interval $(0, n]$ we define two functions f and g by putting

$$f(x) = c_i, \quad g(x) = d_i \quad \text{for } i-1 < x \leq i \quad (i = 1, 2, \dots, n)$$

and then apply to them the integral form of Hölder's inequality given by formula (3) in Chapter X. On computing the integrals entering into this formula we obtain

$$\left| \sum_{i=1}^n c_i d_i \right| \leq \left(\sum_{i=1}^n |c_i|^p \right)^{\frac{1}{p}} \left(\sum_{i=1}^n |d_i|^q \right)^{\frac{1}{q}} \quad (24)$$

This is Hölder's inequality for sums.

Lemma XIII.6.1. *If $F \in A^p$ for certain $p > 1$ the function F is absolutely continuous.*

Proof. First of all we note that if inequality (23) is fulfilled for any partition of the interval $[a, b]$ then for any finite system of nonoverlapping intervals $[a_i, b_i] \subset [a, b]$ ($i = 1, 2, \dots, n$) there must be

$$\sum_{i=1}^n \frac{|F(b_i) - F(a_i)|^p}{(b_i - a_i)^{p-1}} \leq K$$

Now, on the basis of inequality (24), we write

$$\begin{aligned} \sum_{i=1}^n |F(b_i) - F(a_i)| &= \sum_{i=1}^n \frac{|F(b_i) - F(a_i)|}{(b_i - a_i)^{\frac{p-1}{p}}} (b_i - a_i)^{\frac{p-1}{p}} \leq \\ &\leq \left\{ \sum_{i=1}^n \frac{|F(b_i) - F(a_i)|^p}{(b_i - a_i)^{p-1}} \right\}^{\frac{1}{p}} \left\{ \sum_{i=1}^n (b_i - a_i) \right\}^{\frac{1}{q}} \leq \\ &\leq K \left\{ \sum_{i=1}^n (b_i - a_i) \right\}^{\frac{1}{q}} \quad \left(q = \frac{p}{p-1} \right) \end{aligned}$$

Consequently, if $\sum_{i=1}^n (b_i - a_i) < \delta$ then

$$\sum_{i=1}^n |F(b_i) - F(a_i)| < K \delta^{\frac{1}{q}}$$

and δ can be chosen so that this sum becomes arbitrarily small, which proves the absolute continuity of the function F .*

Theorem XIII.6.1. (F. Riesz). *A function F defined on an interval $[a, b]$ is representable by formula (4) with integrand $f \in L^p$ ($p > 1$) if and only if $F \in A^p$.*

Proof. (a) *The "only if" part of the theorem.* Let $f \in L^p$ and F be representable by formula (4). Then, for any two points $x_1, x_2 \in [a, b]$, $x_1 < x_2$, we can write, on the basis of

* We stress that the condition $p > 1$ plays an essential role in his proof.

Hölder's integral inequality, the relation

$$|F(x_2) - F(x_1)|^p = \left| \int_{x_1}^{x_2} f dx \right|^p \leq \left(\int_{x_1}^{x_2} |f|^p dx \right) (x_2 - x_1)^{\frac{p}{q}}$$

where $\frac{1}{p} + \frac{1}{q} = 1$. Hence, for any partition of the interval $[a, b]$ we have

$$\sum_{i=0}^{n-1} \frac{|F(x_{i+1}) - F(x_i)|^p}{(x_{i+1} - x_i)^{p-1}} \leq \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} |f|^p dx = \int_a^b |f|^p dx \quad (25)$$

(b) *The "if" part of the theorem.* Let $F \in \mathcal{A}^p$ ($p > 1$). By the above lemma, the function F is absolutely continuous, and consequently F is representable with the aid of formula (4), and, by Theorem XIII.5.1, $f(x) = F'(x)$ almost everywhere.

For an arbitrary natural number n we put

$$x_i^{(n)} = a + \frac{i}{n}(b-a) \quad (i=0, 1, \dots, n)$$

and construct on the interval $[a, b]$ the piecewise constant functions defined by the formulas

$$f_n(x) = \frac{F(x_{i+1}^{(n)}) - F(x_i^{(n)})}{x_{i+1}^{(n)} - x_i^{(n)}} \quad \text{for } x_i^{(n)} < x < x_{i+1}^{(n)}$$

(at the points $x_i^{(n)}$ the function f_n is not defined). If a point $x \in [a, b]$ does not coincide with any point $x_i^{(n)}$, $n = 1, 2, \dots$, and $f(x) = F'(x)$, we have $f_n(x) \rightarrow f(x)$.*

* If $F'(x_0)$ exists then

$$F'(x_0) = \lim_{\substack{x_1 \rightarrow x_0 - 0 \\ x_2 \rightarrow x_0 + 0}} \frac{F(x_2) - F(x_1)}{x_2 - x_1}$$

Indeed, the obvious equality

$$\begin{aligned} \frac{F(x_2) - F(x_1)}{x_2 - x_1} &= \frac{F(x_2) - F(x_0)}{x_2 - x_0} \cdot \frac{x_2 - x_0}{x_2 - x_1} + \frac{F(x_1) - F(x_0)}{x_1 - x_0} \times \\ &\quad \times \frac{x_0 - x_1}{x_2 - x_1} \end{aligned}$$

implies that the value of the fraction $\frac{F(x_2) - F(x_1)}{x_2 - x_1}$ lies between those of the ratios $\frac{F(x_2) - F(x_0)}{x_2 - x_0}$ and $\frac{F(x_1) - F(x_0)}{x_1 - x_0}$ each of which tends to $F'(x_0)$.

Thus, $f_n(x) \rightarrow f(x)$ almost everywhere on the interval $[a, b]$. Consequently, by Fatou's theorem,

$$\int_a^b |f|^p dx \leq \sup_n \int_a^b |f_n|^p dx$$

But

$$\int_a^b |f_n|^p dx = \sum_{i=0}^{n-1} \frac{|F(x_{i+1}^{(n)}) - F(x_i^{(n)})|^p}{(x_{i+1}^{(n)} - x_i^{(n)})^{p-1}} \leq K$$

and therefore we also have

$$\int_a^b |f|^p dx \leq K \quad (26)$$

whence $f \in L^p$.

Remark. Comparing inequalities (25) and (26) we readily see that

$$\int_a^b |f|^p dx = \sup_{\tau} \sum_{i=0}^{n-1} \frac{|F(x_{i+1}) - F(x_i)|^p}{(x_{i+1} - x_i)^{p-1}}$$

where the supremum is taken over all the possible partitions τ of the interval $[a, b]$. A slight change of the above estimations readily shows that if F is absolutely continuous and $f = F'$ then

$$\int_a^b |f| dx = V_a^b F$$

§ 7. Density Points of a Set on the Line

The techniques used in this chapter enable us to study an important geometrical characteristic of sets on the line.

Definition. Let $E \subset \mathbf{R}_1$ be an arbitrary set (not necessarily a measurable one!) on the line. A point $x_0 \in \mathbf{R}_1$ (which may not belong to E) is called a *density point* of E if

$$\lim_{x_0 \in \Delta, \mu\Delta \rightarrow 0} \frac{\mu^*(E \cap \Delta)}{\mu\Delta} = 1$$

where Δ denotes an arbitrary interval and the limit is taken on condition that Δ contains the point x_0 and contracts to it.

It is evident that every interior point of the set E is its density point.

Theorem XIII.7.1. *Almost all points of the set E are its density points.*

Proof. It is sufficient to consider a set E which is contained within a finite interval (a, b) . Let us construct the function

$$F(x) = \begin{cases} 0 & \text{for } x = a \\ \mu^*(E \cap (a, x)) & \text{for } a < x \leq b \end{cases}$$

This is a continuous increasing function. Moreover, if $x_1 < x_2$ then

$$F(x_2) - F(x_1) = \mu^*(E \cap (x_1, x_2)) \quad (27)$$

This follows from the fact that, since all the intervals are measurable, they "divide well" any set, and therefore

$$\mu^*(E \cap (a, x_2)) = \mu^*(E \cap (a, x_1)) + \mu^*(E \cap (x_1, x_2))$$

(cf., for instance, formula (4) in Chapter IV).

Let us set numbers $\varepsilon_n > 0$ so that $\sum_{n=1}^{\infty} \varepsilon_n < +\infty$ and find, for each n , an open set G_n such that

$$(a) \ E \subset G_n \subset (a, b);$$

$$(b) \ \mu G_n < \mu^* E + \varepsilon_n.$$

Next we put

$$F_n(x) = \begin{cases} 0 & \text{for } x = a \\ \mu(G_n \cap (a, x)) & \text{for } a < x \leq b \end{cases}$$

The functions F_n ($n = 1, 2, \dots$) are also continuous and increasing. Let us show that the differences $\Phi_n = F_n - F$ are increasing functions.

Indeed, if $a \leq x_1 < x_2 \leq b$ then, by virtue of (27),

$$\begin{aligned} \Phi_n(x_2) - \Phi_n(x_1) &= \mu(G_n \cap (x_1, x_2)) - \\ &\quad - \mu^*(E \cap (x_1, x_2)) \geq 0 \end{aligned}$$

We also have

$$0 \leq \Phi_n(x) \leq \Phi_n(b) = \mu G_n - \mu^* E < \varepsilon_n$$

and, consequently, the series $\sum_{n=1}^{\infty} \Phi_n(x)$ is convergent for all $x \in [a, b]$. Therefore, by Theorem XIII.3.3, the series

$\sum_{n=1}^{\infty} \Phi'_n(x)$ is convergent almost everywhere and hence $\Phi'_n(x) \rightarrow 0$ almost everywhere, that is $F'_n(x) \rightarrow F'(x)$ almost everywhere.

From the definition of the function F_n it clearly follows that its increment gained on every component of the open set G_n coincides with the increment of the argument x , and consequently $F'_n(x) = 1$ for $x \in G_n$. Therefore, if $x \in E$, there must be $F'_n(x) = 1$ for all n , and hence $F'(x) = 1$ for almost all $x \in E$.

Let us choose an arbitrary point $x_0 \in E$ at which $F'(x_0) = 1$, and let $\Delta = \langle x_1, x_2 \rangle$ be an interval containing x_0 in its interior: $x_1 < x_0 < x_2$. It follows immediately from formula (27) that

$$\frac{\mu^*(E \cap \Delta)}{\mu\Delta} = \frac{F(x_2) - F(x_1)}{x_2 - x_1} \xrightarrow{x_1, x_2 \rightarrow x_0} F'(x_0) = 1$$

The same result is obtained in the case when x_0 is one of the end-points of the interval Δ (and $\mu\Delta \rightarrow 0$). The theorem has thus been proved.

Remarks. 1°. For a measurable set E the result proved above can be derived in an essentially simpler way from Theorem XIII.5.1 by differentiating the integral of the characteristic function of the set E .

2°. For a measurable set E , considering the density points of its complement, we readily conclude that

$$\lim_{x_0 \in \Delta, \mu\Delta \rightarrow 0} \frac{\mu(E \cap \Delta)}{\mu\Delta} = 0$$

for almost all $x_0 \notin E$. Such points are called *rarefaction points* of the set E .

Thus, the structure of every measurable set on the line is such that almost all the points of the line possess the following property: in the vicinity of each of these points the points of one of the sets E and $\mathbf{R}_1 \setminus E$ are situated "densely" while, conversely, the points of the other set are situated "rarely". There are "few" points in whose vicinity both the points of the set E and the points of its complement are situated "densely"; these points form a set of measure zero. The last assertions are of course of a descriptive character and have no strict mathematical meaning.

The theorem proved implies an interesting consequence: if E is a set on the line having a positive outer measure ($\mu^* E > 0$) there cannot exist a number $\lambda < 1$ such that

$$\mu^* (E \cap \langle a, b \rangle) \leq \lambda (b - a)$$

for any interval $\langle a, b \rangle$. By the way, this conclusion can also be drawn from the proof of Lemma XIII.3.4.

The notion of a density point can be applied to solve in a more general setting the classical problem of differentiation of an integral with respect to its upper limit at the points of continuity of the integrand function. Let f be a summable function on a set $E \subset [a, b]$. For every $x \in [a, b]$ we put

$$F(x) = \int_{E \cap [a, x]} f d\mu$$

In this case, even if $x_0 \in E$ and the function f is continuous at the point x_0 , we can no longer assert (in contrast to the classical case) that $F'(x_0) = f(x_0)$. For instance, if x_0 is an isolated point of the set E then $F'(x_0) = 0$.* However, the following theorem holds.

Theorem XIII.7.2. *If $x_0 \in E$, x_0 is a density point of the set E and the function f is continuous at that point then*

$$F'(x_0) = f(x_0)$$

Proof. We shall show that under the conditions of the theorem the right-hand derivative F'_r of the function F exists at the point x_0 and is equal to $f(x_0)$ ($F'_r(x_0) = f(x_0)$) if $x_0 < b$. We shall confine ourselves to the case of a finite value $f(x_0)$ of the function f at the point x_0 . By the definition of a density point,

$$\frac{1}{h} \mu(E \cap [x_0, x_0 + h]) \xrightarrow{h \rightarrow +0} 1 \quad (28)$$

Setting an arbitrary $\varepsilon > 0$ we can choose $\delta > 0$ such that

$$|f(x) - f(x_0)| < \varepsilon \quad \text{for } x \in E \cap [x_0, x_0 + \delta]$$

* We remind the reader that every function is continuous at any isolated point of the set on which it is defined.

It follows that

$$\begin{aligned} [f(x_0) - \varepsilon] \cdot \mu(E \cap [x_0, x_0 + h]) &< \int_{E \cap [x_0, x_0 + h]} f d\mu < \\ &< [f(x_0) + \varepsilon] \cdot \mu(E \cap [x_0, x_0 + h]) \end{aligned}$$

for $0 < h \leq \delta$, and thus

$$\begin{aligned} [f(x_0) - \varepsilon] \frac{\mu(E \cap [x_0, x_0 + h])}{h} &< \frac{F(x_0 + h) - F(x_0)}{h} < \\ &< [f(x_0) + \varepsilon] \cdot \frac{\mu(E \cap [x_0, x_0 + h])}{h} \end{aligned}$$

Now we see from (28) that

$$f(x_0) - \varepsilon \leq \underline{D}_r F(x_0) \leq \bar{D}_r F(x_0) \leq f(x_0) + \varepsilon$$

By the arbitrariness of ε , this implies

$$\underline{D}_r F(x_0) = \bar{D}_r F(x_0)$$

that is the right-hand derivative exists and

$$F'_r(x_0) = f(x_0)$$

The case when $f(x_0) = \pm \infty$ is treated analogously.

For the left-hand derivative F'_l of the function F the same scheme can be applied to prove that $F'_l(x_0) = f(x_0)$ if $x_0 > a$ and the point x_0 satisfies the conditions of the theorem. Hence, $F'(x_0) = f(x_0)$ for every density point x_0 at which the function f is continuous.

Bibliography*

The present list includes both books devoted primarily to the theory of functions of a real variable and books in which problems related to this theory, especially to the notions of measure and integral, play an essential role.

1. Akilov, G. P., Makarov, B. M., Khavin, V. P. *An Elementary Introduction to the Theory of the Integral*, Leningrad, LGU Publ., 1969 (Russian edition)
2. Aleksandrov, P. S. *Introduction to the General Theory of Sets and Functions*, Moscow, Gostekhizdat, 1948 (Russian edition)
3. Aleksandrov, P. S., Kolmogorov, A. N. *Introduction to the Theory of Functions of a Real Variable*, Moscow, GONTI, 1938 (Russian edition)
4. Bourbaki, N. *Intégration*, Paris, Hermann & Co., Livre VI, Ch. 5, 1956; Ch. 6, 1959
5. Dunford, N., Schwartz, J. *Linear Operators*, New York, Interscience, Part I (*General Theory*), 1958
6. Gikhman, I. I., Skorokhod, A. V. *Introduction to the Theory of Random Processes*, Moscow, "Nauka", 1965 (Russian edition)
7. Halmos, P. S. *Measure Theory*, Princeton, Van Nostrand, 1950
8. Hausdorff, F. *Mengenlehre*, Berlin-Leipzig, W. de Gruyter, 1935
9. Kamke, E. *Das Lebesgue-Stieltjes Integral*, Leipzig, 1956
10. Kolmogorov, A. N., Fomin, S. V. *Elements of the Theory of Functions and Functional Analysis*, Moscow, "Nauka", 1968 (Russian edition). The English translation: *Introductory Real Analysis*, Prentice-Hall, Englewood Cliffs, New Jersey, 1970
11. Lebesgue, H. *Leçons sur l'intégration et la recherche de fonctions primitive*, Paris, 1928
12. Loève, M. *Probability Theory*, Princeton, Van Nostrand, New Jersey, 1963
13. Natanson, I. P. *Theory of Functions of a Real Variable*, Moscow, Gostekhizdat, 1957 (Russian edition). The English translation: *Theory of Functions of a Real Variable*, New York, V. I, 1955; V. 2, 1960
14. Riesz, F., Nagy, Sz. *Leçons d'analyse fonctionnelle*, Budapest, 1953. The English translation: *Functional Analysis*, New York, 1955
15. Saks, S. *Theory of Integral*, New York, Dover Publ., 1966
16. Shilov, G. E. *Mathematical Analysis (Special Course)*, Moscow, Fizmatgiz, 1960 (Russian edition)

* The bibliography includes those English translations of Russian books which are known to the translator.—Tr.

17. Shilov, G. E., Gurevich, B. L. *Integral, Measure and Derivative*, Moscow, "Nauka", 1967 (Russian edition). The English translation: *Integral, Measure and Derivative*, Prentice-Hall, Englewood Cliffs, New Jersey, 1966
18. Smirnov, V. I. *A Course of Higher Mathematics*, V. 5, Moscow, Fizmatgiz, 1959 (Russian edition). The English translation: *A Course of Higher Mathematics*, Reading, Mass., 1964.
19. Sobolev, V. I. *Lectures on Supplementary Topics in Mathematical Analysis*, Moscow, "Nauka", 1968 (Russian edition)
20. Titchmarsh, E. C. *The Theory of Functions*, Oxford University Press, London, 1939

Name Index

Akilov, G. P. 351
Aleksandrov, P. S. 171, 351

Baire, L. R. 126
Banach, S. 74
Bessel, F. W. 248
Bolzano, B. 45
Borel, f.É.É. 53
Budak, B. M. 73
Bunyakovsky, V. Ya. 238
Bourbaki, N. 351

Cantor, G. 11
Carathéodory, G. C. 94
Cauchy, A. L. 41
Cohen, P. 33

Darboux, J. G. 151
Dirichlet, P. G. L. 130
Dunford, N. 351

Egorov, D. F. 141

Fatou, P. 207
Fomin, S. V. 73, 246
Fikhtengolts, G. M. 287
Fischer, E. S. 241
Fourier, J. B. J. 247
Fréchet, R. M. 141
Fubini, G. 229

Gikhman, I. I. 351
Gokhman, E. Kh. 297
Gurevich, B. L. 352

Hahn, H. 313
Halmos, P. R. 313
Hausdorff, F. 351
Hilbert, D. 243
Hildebrandt, T. H. 287
Hölder, O. 260

Jordan, C. 273

Kamke, E. 351
Kantorovich, L. V. 287
Khavin, V. P. 351
Kolmogorov, A. N. 351

Lebesgue, H. 53, 351
Legendre, A. M. 246
Levi, B. 203
Lipschitz, R. O. S. 316
Loéve, M. 351
Luzin, N. N. 141

Makarov, B. M. 10, 351
Minkowski, H. 261

Nagy, Sz. 321, 351
Natanson, G. I. 10
Natanson, I. P. 9, 351
Nikodym, O. 308

Parseval, M. A. 249
Pythagoras 245

Radon, J. 271
Riemann, H. F. B. 83
Riesz, F. 140

Saks, S. 351
Schwartz, J. L. 351
Shilov, G. E. 351
Skorokhod, A. V. 351
Smirnov, V. I. 352
Sobolev, V. I. 352
Steinhaus, H. 220
Stieltjes, T. J. 284

Titchmarsh, E. C. 352
Tonelli, L. 230

Uryson, P. S. 171

Vitali, G. 149
Vladimirov, D. A. 10
Vulikh, B. Z. 74

Weierstrass, K. T. W. 45

Yosida, K. 74

Subject Index

- Absolute continuity of the integral 192
- Absolute equicontinuity of integrals 209
- Algebra 35
 - generated by a set 37
- Almost everywhere 132
- Bessel's inequality 248
- Bunyakovsky's inequality 238
- Cauchy sequence 69
- Cauchy's inequality 41, 61
- Cell, n -dimensional 109
- Closed sphere 66
- Closure 68
- Complement of a set 16
- Component of an open set on the line 49
- Convergence
 - almost everywhere 132
 - in the mean of order 1 214
 - in the mean of order p 262
 - in the mean of order 2 240
 - in measure 135
 - in metric space 63
 - in norm 76
- (*)-convergence 171
- Convergence regulator 143
- Countable semi-additivity of a measure (of outer measure) 88
- Covering 53
- Derived numbers 323
- Difference of sets 14
- Disjoint sets 17
- Distance 59
- Distribution function 299
- Distributive laws for unions and intersections of sets 14
- Equality of sets 12
- Equivalence relation 172
- Equivalent functions 133
- Equivalent sets 19
- Essential supremum, 219
- Factor set 173
- Fourier's
 - coefficients 247
 - series 247
- Function
 - absolutely continuous 306
 - of Baire class 1 130
 - of Baire class 2 130
 - bounded almost everywhere 219
 - of bounded variation 280
 - Lebesgue integrable 152
 - Lebesgue measurable 122
 - measurable 122
 - square summable 237
 - summable 180, 181
 - $\mathfrak{B}$ -measurable (Baire) 126
 - $\mathfrak{C}$ -measurable 122
- Functional 78
 - additive 79
 - bounded 81
 - continuous 78
 - distributive 79
 - homogeneous 79
 - linear 79
- Gluing principle 21
- Hahn decomposition 313
- Hölder's inequality
 - for integrals 260
 - for sums 343
- Image of an element 19
- Integral
 - improper 199
 - Lebesgue's, of a bounded function 152
 - Lebesgue's, of an unbounded function 180ff
 - Lebesgue-Stieltjes 284

- Integral (*cont.*)
 Radon's 283
 Stieltjes' 294
 Intersection of sets 14
 Isometric spaces 253
 Isomorphic spaces 253

 Jordan decomposition 273
 Jumps of a monotone function 331

 Lebesgue-Darboux sums 152
 Limit of a sequence of points 42, 63
 Linear function space 74
 Lipschitz condition 316

 Mean value of a function 262
 weighted 263
 Measure 86
 complete 95
 finite 86
 generated by outer measure 93
 Lebesgue's, in a Euclidean space 115
 outer 88
 σ -finite 86
 Metrization 63
 Minkowski's inequality 261

 Neighbourhood of a point 43, 66
 Network partition of a parallelepiped 104
 Norm
 of an element 75
 of a functional 81

 One-to-one correspondence 18
 Open sphere 43, 66
 Orthogonal functions 244
 Orthogonal system of functions 246
 Orthonormal system of functions 247
 complete 249
 Outer measure 88
 generated by a measure 90

 Lebesgue's, in a Euclidean space 115

 Parallelepiped
 closed 101
 n -dimensional 101
 open 101
 with finite edges 102
 Parseval's relation 249
 Point
 accumulation 43
 density 346
 interior 68
 isolated 44
 Lebesgue 341
 limit 43, 66
 of an n -dimensional space 39
 rarefaction 348
 Point function 315
 Power
 of a set 27
 Product measure 235
 Product of sets 233
 Projection of a point 221

 Radon-Nikodym derivative 331
 Ring 34
 generated by a set 37

 Scalar product of functions 242
 Semiring 37
 Set(s)
 Borel 58
 bounded, in a Euclidean space 44
 bounded, in a metric space 69
 bounded, in a normed space 77
 closed 46, 66
 countable 21
 empty 13
 everywhere dense 72
 F_σ 58
 G_δ 58
 Lebesgue's, of a function 121
 Lebesgue measurable, in a

- Euclidean space 115
 - on the line 49ff
 - measurable 92
 - m -measurable 96
 - open 49, 66
 - of the same power 19
 - uncountable 25
 - μ^* -measurable 91
- Set function 83
 - continuous 55
 - finitely additive 83
 - σ -additive 83
- Space
 - Banach 77
 - Euclidean, infinite-dimensional 61
 - Hilbert 243
 - n -dimensional 40
 - metric 59
 - complete 71
 - separable 72
- Stability of convergence 141
- Standard extension of measure 96
- Stieltjes-Darboux sums 294
- Subgraph 221
- Subset 12
- Theorem
 - Bözano-Weierstrass 45, 46
 - Borel-Lebesgue 53
 - on a common convergence regulator 174
 - on diagonal sequence 175
- Egorov's 143
- Fatou's 207
- Fréchet's 148
- Fubini's, on iterated integrals 229
- Fubini's, on term-by-term differentiation of series 329
- Lebesgue's, on differentiation of a monotone function 327
- Lebesgue's, on passage to the limit under the integral sign 165, 200
- Levi's 203
- Luzin's 144
- Radon-Nikodym 308
- Tonelli's 230
- Truncated function 204
- Union of sets 13
- Variation
 - negative 271
 - positive 271
 - total 271
- Volume of a parallelepiped 104
- Zero element 75
- σ -algebra 36
- σ -ring 35

TO THE READER

Mir Publishers would be grateful for your comments on the content, translation and design of this book. We would also be pleased to receive any other suggestions you may wish to make.

Our address is:
USSR, 129820, Moscow I-110, GSP
Pervy Rizhsky Pereulok, 2
MIR PUBLISHERS

Printed in the Union of Soviet Socialist Republics

Other Books by Mir Publishers

Minorsky, V. P.

PROBLEMS IN HIGHER MATHEMATICS

396 pp.

A collection of 2570 problems in analytical geometry and mathematical analysis for university engineering students. As each section opens with the formulas, definitions and other theory needed for solution of the problems that follow, and answers and indications on how to go about the solution are provided for many of the problems, the book can be used either under the supervision of an instructor or for independent study by correspondence and extramural students. Each section closes with questions for revision, this material constituting one-third of the volume of the book.

Budak, B. M. and Fomin, S .V.

MULTIPLE INTEGRALS, FIELD THEORY AND SERIES

640 pp.

Covers branches of mathematics increasingly required by physicists, such as multiple, line and improper integrals, the theory of fields, and power and trigonometric series. Based on lectures read by the authors in the physics faculty of Moscow University, the book endeavours to show the connection between the various mathematical concepts and their applications, and wherever possible their physical sense as well.

